THE UNIVERSITY
of EDINBURGH

De novo Biological Engineering of a tRNA Neochromosome in Yeast

Roy Scott Kamla Walker

This thesis is submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

At the:

Institute *for* Bioengineering
School of Engineering
University of Edinburgh

2017

THE UNIVERSITY
of EDINBURGH

Abstract

Advances in DNA synthesis technology have led to rapid growth in the field of synthetic biology, heralding a nascent era of synthetic genomics. Sc2.0 (*Saccharomyces cerevisiae* version 2.0) is an international consortium with the aim of designing and constructing a fully-synthetic eukaryotic genome. Fundamental design changes to the synthetic genome include the removal of unstable tRNA genes and their intended collation onto a "tRNA neochromosome", with the aim of producing a more robust and stable synthetic genome structure. To maintain viability of a synthetic yeast, the tRNA neochromosome is therefore considered an important if not essential aspect of this project.

The application of engineering principles is synonymous with synthetic biology, regularly employing the recursive Design-Build-Test cycle to improve experimental approach. This doctoral study explores the design, construction and characterisation of a tRNA neochromosome in *Saccharomyces cerevisiae*. A series of design principles influenced by engineering concepts were used to rationalise the complexities of *de novo* chromosome engineering, maximise its stability and ensure function *in vivo*. A methodology based on *in vivo* homologous recombination was then developed and shown to reliably construct the neochromosome from its constituent parts. Experimental characterisation revealed that genetic elements function as expected, and that the parental strain can tolerate the sole presence of one each of three single-copy, essential tRNA genes (*SUP61*, *TRT2* and *TRR4*), although Northern blot revealed potential precursor accumulation of the SUP61 tRNA caused by the presence of a synthetic 5' flanking sequence. Following the addition of synthetic telomere seed sequences, pulsed-field gel electrophoresis (PFGE) and deep sequencing revealed complex structure variations in two independent strain backgrounds. Except for these structural variations, successful neochromosome construction demonstrated the applicability of the approaches used and the remarkable ability of the yeast model to support the presence of a 17th chromosome housing an additional 275 tRNA genes. The research in this thesis has for the first time described the design, construction and characterisation of a eukaryotic neochromosome *de novo*. It is hoped that the findings presented will further our understanding of tRNA biology and enhance the aims of the Sc2.0 project.

Lay Summary

Advances in DNA synthesis technology have led to rapid growth in the field of synthetic biology, heralding a nascent era of synthetic genomics. Sc2.0 (*Saccharomyces cerevisiae* version 2.0) is an international consortium with the overall aim of designing and constructing a fully-synthetic genome of the humble brewer's yeast. As part of Sc2.0, fundamental design changes include the removal of unstable tRNA genes and their intended collation onto a "tRNA neochromosome", with the intended aim of producing a more robust and stable synthetic genome structure. To maintain viability of a synthetic cell, the tRNA neochromosome is therefore considered an important if not essential aspect of this project.

The application of engineering principles is synonymous with synthetic biology, regularly utilising the recursive design-build-test cycle to rationalise experimental approach. The design, construction and characterisation of a tRNA neochromosome is the primary focus of this doctoral study.

A design-based approach was used to rationalise the complexities of *de novo* chromosome engineering. A methodology was developed and shown to reliably construct the neochromosome from its constituent parts. Experimental characterisation reveals that genetic elements generally function as expected, although visualisation and deep sequencing reveal some structural variations. Overall, successful construction demonstrates the feasibility of the approaches used and the remarkable ability of the yeast model to support the presence of an entire 17th chromosome housing an additional 275 tRNA genes.

Dedication

To Pia, Danica and Graeme

Yours Aye,

Roy

Acknowledgements

[This work was supported by an EPSRC studentship, with a continuation in funding supported by the Bill & Melinda Gates Foundation]

I would like to display appreciation to the following individuals:
Firstly, I would like to thank Dr. Patrick Cai for the opportunity to work on the neochromosome project, your invaluable guidance throughout my PhD, financial support and the continual push for us all to publish. I also owe sincere gratitude to Prof. Alistair Elfick, whose support and guidance in difficult times has inspired me to pursue this challenging, yet fascinating, project. Additional thanks to Alistair, but also to Erika Szymanski and Daniel Schindler for your feedback on my thesis.

Many thanks to all at the Cai lab, including Chantal, Jamie, Darek, Wei, Paulina, Andrea, Aileen, Emily, Katarina, Eva, Erika and Daniel, who have displayed good humour and support throughout these past few years. Specific thanks for Katarina for your invaluable help and advice (my stipend has at least benefited the continual exchange of muffins) as well as Darek for your humour and invaluable help with Northern blot. Additional thanks to those at the Edinburgh Genome Foundry, including Anaïs whose interactions for various art festivals I have found to be rewarding experiences.

Without extensive collaborative efforts, the full potential of the neochromosome project would not have been realised. Specific thanks to Prof. Jef Boeke: the originator, instigator and grandfather of the Sc2.0 consortium. Sincere gratitude to everyone else, including Prof. David Tollervey for the use of your pulsed-field apparatus as well as Tomek for your expertise in tRNA biology and CRAC. Additional gratitude goes to Conrad Nieduszynski, Aaron Brooks, Daniel Schraivogel, Romain Koszul, Guillaume Mercy and Francisco Antequera. I would also like to thank Prof. Adele Marston for your invaluable comments on my proposed model for neochromosome instability.

I would like to thank my supportive family (Pia, Danica and Graeme), who have always provided a home and a regular supply of beer to lubricate the wheels of my journey. I would also like to express gratitude to my father, Graeme, who has inspired me to study the amazing world of yeast (yet supported me without pressure in whatever direction I choose to take).

Lastly, I would like to thank Forbes, whose truism I believe defines the efforts of all PhD students:

"It always takes longer than you think"

Dr. Forbes Wardrop

9 May 1972 to 5th April 2015

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Roy Scott Kamla Walker

Abbreviations

%	-	Percent
5-FOA	-	5-Fluoroorotic acid
A. gossypii	-	*Eremothecium* (*Ashbya) gossypii*
Amp	-	Ampicillin
ARS	-	Autonomously Replicating Sequence
ATP	-	Adenosine tri-phosphate
[γ^{32}P]-ATP	-	Adenosine triphosphate (labelled on the gamma phosphate group with ^{32}P)
bp	-	Base pairs
^{o}C	-	Degrees Celsius
$CaCl_2$	-	Calcium chloride
CEN	-	Centromere
Chr	-	Chromosome
cm	-	Centimetre
ddH_2O	-	Double-distilled water
DEPC	-	Diethyl pyrocarbonate
DMSO	-	Dimethyl sulphoxide
DNA	-	Deoxyribonucleic acid
dNTPs	-	Deoxynucleotide triphosphates
DTT	-	Dithiothreitol
E. cymbalariae	-	*Eremothecium cymbalariae*
E. coli	-	*Escherichia coli*
EDTA	-	Ethylenediaminetetra-acetic acid
g	-	Grams (when referring to weight)
g	-	G-force (when referring to centrifuge speed)
G418	-	Geneticin
His	-	Histidine

kb	-	Kilobase pairs
Kan	-	Kanamycin
L	-	Litre
LB	-	Luria-Bertani media
Leu	-	Leucine
LiOAc	-	Lithium acetate
LTR	-	Long Terminal Repeat
M	-	Moles
mg	-	Milligram
$MgSO_4$	-	Magnesium sulphate
min	-	Minutes
mL	-	Mililitre
mM	-	Millimolar
$MnCl_2$	-	Manganese chloride
MOPS	-	3-(N-morpholino)propanesulfonic acid buffer
mRNA	-	Messenger ribonucleic acid
NAD	-	Nicotinamide adenine dinucleotide
NaOH	-	Sodium hydroxide
OD_{600nm}	-	Optical density (600 nanometres)
PBS	-	Phosphate Buffered Saline
PAGE	-	Polyacrylamide gel electrophoresis
PEG	-	Polyethylene glycol
PCR	-	Polymerase Chain Reaction
PFGE	-	Pulsed Field Gel Electrophoresis
pol II	-	RNA Polymerase II
pol III	-	RNA Polymerase III
RbCl	-	Rubidium chloride
RNA	-	Ribonucleic acid
rRNA	-	Ribosomal ribonucleic acid

rDNA	-	Ribosomal deoxyribonucleic acid
s	-	Seconds
S. cerevisiae	-	*Saccharomyces cerevisiae*
SC	-	Synthetic complete media
SCRaMbLE	-	Synthetic Chromosome Recombination and Modification by LoxP-mediated Evolution
SDS	-	Sodium dodecyl sulphate
Syn	-	Synthetic
SynIII/VI/IXR	-	Synthetic chromosome III, VI and the right arm of IX
TAE	-	Tris-HCl, acetic acid, EDTA
TBE	-	Tris-HCl, boric acid, EDTA
tDNA	-	Transfer deoxyribonucleic acid
tRNA	-	Transfer ribonucleic acid
Tris-HCl	-	Tris-buffered hydrochloric acid
µg	-	Microgram
UHR	-	Universal Homologous Region
µL	-	Microlitre
Ura	-	Uracil
V	-	Volts
v/v	-	volume/volume
WT	-	Wild-type
X	-	Times
YPD	-	Yeast extract, peptone, dextrose

Nomenclature

The following is a summary of the nomenclature and naming standards used throughout this thesis.

- Synthetic, dual-block replication termination cassettes are referred to in italics (*i.e*. *Ter4*).
- Origins of replication as a function are non-italicised (*i.e.* ARS elements), and are italicised when referring to a specific origin of replication (*i.e. chrF-444*).
- Throughout this thesis, tRNA genes are italicised and the tRNA gene products are non-italicised. tRNA genes and their gene products may be referred to in multiple ways, such as by their common name (*e.g. SUP61*), their isotype ($tRNA^{Ser}$ or tS), their anticodon ($tRNA^{Ser}$(CGA)) or their systematic name (e.g. tS(CGA)C). A naming convention therefore appears to differ throughout the literature. For example, Turowski *et al.* (2016) refers to tRNA genes as their non-italicised systematic name while Kutter *et al.* (2011) refer to tRNA genes as their non-italicised anticodon name and Copela *et al.* (2006) refer to tRNA molecules by their anticodon name. To simplify naming and prevent confusion with the systematic name (*i.e.* the systematic name for *SUP61* is tS(CGA)C), tRNA genes are referred to as their common (or standard) name in italics (*i.e. SUP61*), and references to the tRNA molecule itself are non-italicised (*i.e.* SUP61).
- The *HO* locus is referred to in italics (*i.e*. *HO* locus).
- All other genes names are italicised (*i.e. KanMX*).

Table of Contents

Index of Figures

Index of Tables

Chapter 1: Introduction

1.1. A Brief History of Synthetic Genomics

Engineering on the scale of the genome, only recently rendered feasible through rapid advances in synthetic biology, now presents the opportunity to ask fundamental questions of biological significance. Indeed, the nascent field of synthetic genomics will come to act as a key enabler in the years to come, proving a means to study biology in a systematic manner and shed new insight into the underlying mechanisms of life. Synthetic genomics also presents the opportunity to harness the full potential of biology for the production of drugs, biofuels, chemicals and vaccines (König *et al.*, 2013).

These radical new approaches to biology would not have existed without the rapid technological advances observed over recent decades. In fact, the origins of synthetic biology can be traced back up to fifty years earlier than its current terminology: in the 1970s' and 1980s', genetic manipulation through cloning and recombinant gene expression was widespread. The first example of genetic modification through restriction cloning was described in the early 1970's, when Cohen *et al*. (1973) successfully cloned a kanamycin resistance marker into the pSC101 plasmid. However, the first true example of *de novo* gene synthesis was performed three years earlier, when Agarwal *et al.* (1970) chemically constructed a tRNA molecule from its constituent polynucleotides. This endeavour took considerable time and effort; today, gene synthesis is performed on a scale orders-of-magnitude greater. Indeed, this thesis describes the *de novo* synthesis and construction of a neochromosome housing 275 tRNA genes.

In the 1990's, DNA sequencing technologies ('Read') have led to a revolution in our understanding of the biological functions encoded in genetic material, and represents a major pillar supporting the efforts of today's synthetic biology. A reduction in cost and an

increase in the speed of sequencing at the rate of Moore's law (Pettersson *et al*., 2009) have allowed this technology to become widespread, ultimately leading to genome-scale sequencing projects. It is pertinent to note that the first eukaryotic organism sequenced was the model, *Saccharomyces cerevisiae* (*S. cerevisiae*) (Goffeau *et al*., 1996).

More recent advances in DNA synthesis technologies ('Write'), have provided a second major pillar supporting the development of synthetic biology. As observed with DNA sequencing, DNA synthesis technologies have also displayed a Moore's law-like trajectory in terms of increased speed and reduced cost. The ability to produce DNA sequences on demand now presents biologists with an unprecedented degree of control over the building blocks of life.

DNA synthesis projects have also increased in scale: Gibson *et al.* (2010) successfully synthesised a 1.08 Mb version of the *Mycoplasma mycoides* genome and subsequently transplanted this DNA into a *Mycoplasma capricolum* host cell. This work represented the first time an entire genome has been designed on a computer, chemically synthesised and shown to successfully replicate inside a cell. The Sc2.0 consortium (Dymond *et al.,* 2011) only initiated one year later, and represents a natural progression in synthetic genome engineering from the prokaryotic domain to the re-design and chemical synthesis of the first eukaryote, *S. cerevisiae*. Five years later, and the scale of synthetic genome engineering increased still: The Genome Project-Write (GP-Write; formerly the *Human* Genome Project-Write) proposes to perform whole-genome engineering on mammalian cell lines and massively reduce the cost of doing so (Boeke *et al.*, 2016, Walker and Cai, 2016).

Each of the ambitious projects described above would not be possible without the development of numerous toolkits and individual enabling technologies. These include the aforementioned rapid advances in DNA sequencing and synthesis technologies, but

also include comparatively recent abilities to manipulate and join DNA on the molecular level. Technologies such as Gibson Assembly (Gibson *et al.*, 2009), CRISPR-Cas9 (Garneau *et al.*, 2010) and Golden-Gate assembly (Engler *et al.*, 2008) have provided synthetic biologists with unprecedented flexibility to construct and alter DNA sequences at will. Powerful software packages, such as Snapgene (Chicago, USA), have greatly improved the workflow and speed of DNA sequence manipulation.

Perhaps more significant to the development of synthetic biology is the application of engineering principles that are now synonymous with this field. These engineering approaches are discussed in further detail and applied throughout this thesis, but include concepts such as the use of standardised biological parts, orthogonality and the Design-Build-Test cycle.

Although several definitions exist (Osbourn *et al*., 2012) synthetic biology may be described, in the broadest sense, as simply *the engineering of biology*.

1.2. The Synthetic Yeast (Sc2.0) Consortium

The Synthetic Yeast (Sc2.0) consortium is an international collaboration that aims to design and synthesise a fully-synthetic eukaryotic genome. This ground-breaking project has led to the re-design and complete chemical synthesis of *S. cerevisiae* chromosome II (Shen *et al.*, 2017), chromosome III (Annaluru *et al*., 2014), chromosome V (Xie *et al.*, 2017), chromosome VI (Mitchell *et al.*, 2017), chromosome X (Wu *et al.*, 2017) and chromosome XII (Zhang *et al.*, 2017). These efforts have helped to shed new insight into the structure and function of the rDNA locus (SynXII) as well as the genome structure of synthetic chromosomes (Mercy *et al*., 2017), comprehensive 'omic analysis of a synthetic chromosome (SynII), new debugging techniques (SynX, SynVI), the generation of poly-synthetic-chromosome strains (Mitchell *et al.,* 2017) and, finally, allowing undergraduate

students to contribute to the synthetic yeast consortium through the innovative 'Build a Genome' course (SynV, SynIII; Cooper *et al.* (2012)).

Sc2.0 aims to provide new insight into the fundamental properties of eukaryotic biology through a 'build to understand' approach, and address questions on diverse topics such as the role of RNA splicing, tRNA biology and the structure of the yeast genome (Dymond et al., 2011, Annaluru et al., 2014, Richardson et al., 2017). Sc2.0 also seeks to develop enabling technologies to advance the field of synthetic genomics and to provide a chassis to develop gain-of-function in synthetic yeast strains through the innovative SCRaMbLE system (described in further detail below).

A defining feature of the Sc2.0 project is the introduction of fundamental designer changes. These alterations to the synthetic genome include the introduction of *loxPsym* sites for systematic combinatorial mutagenesis, PCR tags for verification of synthetic DNA incorporation, stop codons replaced from TAG to TAA to allow future incorporation of artificial amino acids, and the removal of retrotransposons, subtelomeric repeats and many introns (Dymond *et al.*, 2011, Annaluru *et al.*, 2014). Perhaps one of the more radical design changes to the synthetic yeast genome is the removal of unstable tRNA genes and their intended relocation onto a tRNA neochromosome (discussed in further detail in Section 1.6). Major design alterations to the synthetic yeast genome sequence are summarised in **Table 1.1**.

Table 1.1: Summary of major design changes undertaken to the Sc2.0 genome sequence. These design changes are described in (Richardson et al., 2017).

Insert	Remove	Relocate	Change
PCR tags	Retrotransposons	tRNA genes to tRNA Neochromosome	TAG stop codons to TAA
loxPsym sites	High-copy-number genes Most introns (including tRNA gene introns) Repetitive elements (including subtelomeric repeats and homopolymer tracts)		Native telomeres to synthetic telomeres

SCRaMbLE (or Synthetic Chromosome Recombination and Modification by *LoxP*-mediated Evolution) is significant technology for the Sc2.0 consortium. Scattered throughout the synthetic yeast genome lie symmetrical *loxPsym* sites located in the 3' region of non-essential genes and at important landmarks (Dymond *et al.*, 2011). These are targets for the site-specific Cre recombinase enzyme, and present the opportunity to systematically alter the structure and content of the synthetic yeast genome by inducing gene deletion, inversion, duplication and translocation. The function of the site-specific Cre recombinase may also be tightly controlled by fusing the protein to a domain induced by micromolar quantities of estradiol (Cre-EBD) (Lindstrom and Gottschling, 2009).

Preliminary work performed on the circular right arm of synthetic chromosome IX (synIXR) reveals extensive structural diversity of SCRaMbLE'd strains (Shen *et al.*, 2016). A key feature of this technology is simply to allow the SCRaMbLE'd yeast to identify the favoured genome arrangement for optimal growth under different conditions. Finally, this technology also introduces the potential to systematically incorporate new metabolic pathways, presenting a unique method to develop gain-of-function in synthetic strains.

A strategy called SwAP-In (switching auxotrophies progressively for integration) was developed to construct synthetic yeast chromosomes (Richardson et al., 2017). This

approach utilises the great power of *in vivo* homologous recombination in yeast to segmentally swap existing wild-type chromosomal DNA with its synthetic counterpart. Shorter synthetic 10 kb chunks are first assembled into megachunks (30 to 60 kb long) by restriction digest ligation followed by stepwise integration into the yeast genome through progressive *LEU2*/*URA3* selective marker swapping. A similar method was developed to construct the tRNA neochromosome (described in Chapter 4)

Of all questions that may be posited: why should scientists build a synthetic version of the humble brewer's yeast? The answers are numerous, but include this organism's past and potential value to science and industry and its well-characterised genome and physiology. *S. cerevisiae* is a single-celled fungus domesticated since ancient times to produce bread and alcoholic beverages. This organism has more recently become an important model organism for laboratory research, displaying the advantages of rapid growth, ease of genetic transformation and a well-understood physiology. Industrially, yeast is involved in the production of biofuels, brewing, baking, flavour additives (Stewart and Priest, 2006) and the bioremediation of heavy metals (Soares and Soares, 2012). A major advantage of constructing a synthetic version of *S. cerevisiae* lies with the relatively small, but extensively sequenced and well-characterised genome. The yeast genome is also of additional use in medicine due to the high degree of conservation in evolution with that of the human genome (Tugendreich *et al.,* 1994).

For synthetic biology, the power and efficiency of the homologous recombination machinery in *S. cerevisiae* has provided a way of combining large segments of DNA. For example, the 592 kb synthetic *Mycoplasma genitalium* genome was assembled in yeast by combining a series of 25 overlapping DNA fragments in a single transformation (Gibson *et al.*, 2008). The homologous recombination machinery has also allowed members of the Sc2.0 consortium to replace existing yeast genomic DNA with its designer counterpart through the SwAP-In strategy, presenting the opportunity to completely alter the yeast

genome according to user-defined design principles. This new organism, *Saccharomyces cerevisiae* version 2.0, has recently been described as *Syntheomyces cerevisiae* (Abbas and Nielsen, 2016).

1.3. Overview of tRNA Biology, Genetics and Contribution to Genomic Instability

tRNA genetics and biology are important topics to fully explain the underlying design decisions undertaken for the tRNA neochromosome. This section presents an overview of tRNA biology, including transcription, processing and splicing, with an overall focus on the yeast model since much of the current understanding of tRNA biology originates from their study in *S. cerevisiae*. The latter part of this section describes the contribution of tRNA genes to replication stress and genome instability, important to explain the rationale for tRNA gene relocation from the synthetic yeast genome.

1.3.1. tRNA Biology: An Overview

The central dogma of molecular biology (Crick, 1970) describes the transfer of information in biological systems. Fundamentally, this involves the transfer of information encoded in DNA to the biological machinery that converts this information into protein. This process is largely separated into two steps: transcription and translation.

Transfer RNA (tRNA) molecules are critical to the process of translation. These small, folded stretches of ribonucleic acid (**Figure 1.1**) serve as adaptors that link the information found in the three-base codons of mRNA to the polypeptide chains that form proteins. Due to their central role in biology, tRNAs are evolutionarily ancient and essential to all known forms of life (Sun and Caetano-Anolles, 2008).

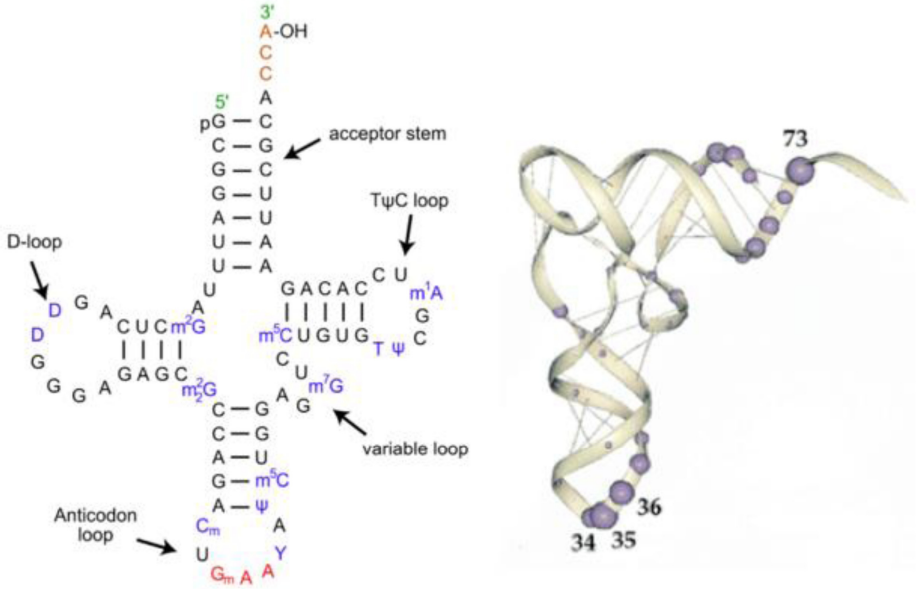


Figure 1.1: Standard "cloverleaf" secondary structure of tRNA molecules. The image on the left represents an annotated two-dimensional tRNA form and the image on the right represents its folded three-dimensional structure. On the left image, the anticodon may be observed as three red bases in the anticodon loop and base modifications are represented by nucleotides in purple. The orange CCA tail is the location of amino acid binding to respective tRNA molecules. The above image was adapted from Hartman and Smith (2014).

As transcription is the foundation of mRNA synthesis from DNA, translation is the foundation of protein biosynthesis. tRNA molecules assist in the decoding of the three-base codon of mRNA through complementary matching of the tRNA anticodon. The ribosome mediates codon-anticodon matching, and catalyses the transfer of each amino acid specific to each tRNA molecule into a growing polypeptide chain.

As mRNA codons in the standard genetic code are read in triplicate bases, a total of 64 codons may exist (i.e. 4^3 possible combinations of U, A, G and C). Once the three stop codons are considered, a total of 61 individual tRNA species would be required to align with all possible mRNA codons. However, cells generally contain significantly fewer tRNA species, with *S. cerevisiae* containing 42 distinct isoacceptor classes (Percudani *et al.*,

1997). Therefore, tRNA molecules must match with more than one codon to assign the 20 standard amino acids to the standard genetic code.

Codon degeneracy (or redundancy) refers to the propensity for some tRNA molecules to pair with more than one codon, mediated by a modification on the first anticodon base of several tRNAs. This 'wobble base pair' enables the translation of all possible codons found in the genetic code. Additionally, as 42 isoacceptor classes are required to generate the 20 standard amino acids, different tRNA molecules may encode the same amino acid. These are referred to as isoacceptor families.

Due to the redundancy of the genetic code, some codons are preferred over others in different organisms. This bias is also inherent in the copy number of tRNA genes required to decode these codons: a strong correlation exists between tRNA gene copy number and usage in translation. Overall, this is referred to as codon usage bias (Percudani *et al.*, 1997).

1.3.2. tRNA Transcription and Regulation

tRNA genes in eukaryotes are transcribed by the enzyme, RNA Polymerase III (Pol III). Recruitment of this multi-part enzyme complex requires two transcription factors: TFIIIB and TFIIIC. Transcription is initiated when TFIIIC binds to a dual conserved promoter element (A and B sequence blocks) internal to the tRNA gene itself, permitting recruitment of TFIIIB and subsequent recruitment of Pol III (Schramm and Hernandez, 2002). Transcriptional termination is mediated by a stretch of poly-thymidine residues, with five or more encountered in the genome of *S. cerevisiae* (Braglia *et al.,* 2005). Pol III is negatively regulated in a global manner by the protein, Maf1, which is conserved across all known eukaryotes (Pluta *et al.*, 2001, Karkusiewicz *et al.,* 2011). Under conditions that render it advantageous to reduce translational activity, such as that of starvation and

DNA damage, Maf1 is dephosphorylated and imported into the nucleus where it inhibits transcription by binding Pol III. Under favourable growth conditions, Maf1 is phosphorylated and is transported back into the cytoplasm so tRNA gene expression may occur unhindered.

1.3.3. tRNA Processing and Modification

The progression from a precursor tRNA to the mature final form in yeast requires five major steps (Hopper and Phizicky, 2003). These include removal of the 5' sequence, removal of the 3' trailer sequence, addition of a trinucleotide CCA for subsequent amino acid charging, splicing of tRNA introns and numerous base modifications.

The 5' end of each precursor tRNA molecule is processed by the highly-unusual and ancient ribozyme endonuclease, RNase P (Bertrand *et al.*, 1998, Chamberlain *et al.*, 1998). This enzyme complex is the largest in the tRNA processing pathway and consists of nine essential protein sub-units and, interestingly, one essential RNA subunit that is responsible for its catalytic activity. The processing of the 3' end of precursor tRNAs is understood to a lesser degree. However, this complex process includes both endonuclease and exonuclease activity: the endonuclease, RNaseZ, or Trz1, appears to play the major role in 3' processing (Chen *et al.*, 2005, Maraia and Lamichhane, 2011) although the exonuclease Rex1 plays a major role in 3' to 5' processing activity (Copela *et al.*, 2008). The La protein is responsible for stabilising the 3' region for endonuclease activity to occur and protects this region against premature exonuclease activity (Yoo and Wolin, 1997). Following 3' processing, the addition of the trinucleotide CCA onto the 3' terminus is performed by the enzyme tRNA nucleotidyl transferase, subsequently essential for charging each tRNA with its complementary amino acid (Vortler and Morl, 2010).

1.3.4. Splicing of tRNA Introns

tRNA introns are found in all three domains of life, and are likely evolutionarily ancient (Kanai, 2013), although it is interesting to note that yeast mitochondrial tRNAs do not contain introns (Chen and Martin, 1988). tRNA splicing is an essential process throughout all eukaryotes (Dhungel and Hopper, 2012) and *S. cerevisiae* is no exception: around 20% of all tRNA genes in this organism contain introns (Hopper, 2013). tRNA splicing in yeast is undertaken by three primary enzymes in three distinct steps that include splicing, ligation and removal of a residual phosphate. tRNA splicing endonuclease, responsible for excising the tRNA intron, is a four-subunit complex consisting of Sen2, Sen15, Sen34 and Sen54 (Trotta *et al.*, 1997). All four subunits are essential. tRNA ligase (Trl1) and 2'phosphotransferase (Tpt1) are responsible for ligation of the two tRNA exons and removal of a residual phosphate, respectively (Phizicky et al., 1986, Culver *et al*., 1997).

Early work in other eukaryotes suggested that tRNA splicing occurs solely in the nucleus (Hopper *et al.,* 2010, Clark and Abelson, 1987). However, Yoshihisa *et al.* (2003) show that tRNA splicing in yeast is unexpectedly undertaken on the cytoplasmic side of the mitochondria. The reasons for this are unclear, although it has been proposed that the SEN complex plays a moonlighting role in rRNA processing (Dhungel and Hopper, 2012).

Overall, the general function of tRNA introns is unknown. Some evidence suggests that they play a role in pseudouridine modification of the anticodon (Szweykowska-Kulinska *et al.,* 1994), although it should be noted that the presence of introns in other tRNAs are not necessarily correlated with modification of the anticodon (Johnson and Abelson, 1983). It should also be noted that, at least in the case of $tRNA^{Ile}$, anticodon modification is not dependent on the mechanism of splicing itself (Szweykowska-Kulinska *et al.*, 1994) i.e. modification of the tRNA species occurs before intron excision itself.

The essentiality of tRNA introns remains very much an open question. Mori *et al.* (2011) performed a systematic study by removing all introns from all six copies of the $tRNA^{Trp}$ gene. While all copies of this tRNA are essential, interestingly, no significant defects were observed after their removal and $tRNA^{Trp}$ expression levels were observed to be normal.

1.4. tRNA Genes and Their Contribution to Genomic Instability

The central motivation of the tRNA neochromosome project is to introduce greater stability in the synthetic yeast genome by relocating unstable tRNA genes onto a dedicated chromosome. This section presents an overview of current literature describing the mechanistic nature of genomic instability associated with tRNA genes. Direct sources of instability include 'collisions' between the replication fork and components associated with transcription, leading to the formation of DNA lesions and subsequent repair through recombination. It is well known that actively-transcribed tRNA genes are 'replication-fork barriers', although the exact mechanism of replication fork pausing and/or collapse at tRNA genes is unknown (Nguyen *et al.*, 2010). The consequences of replication stress are discussed in greater detail in this section due to its greater overall contribution to genomic instability. Indirect sources of genome instability include incorporation of retrotransposons and their associated repeats at regions proximal to tRNA genes. Finally, as the nature of genomic instability is conserved throughout eukaryotes, the mechanisms of replications stress from mammalian cells to yeast, including common features associated with genes transcribed by RNA polymerase II, is discussed below.

1.4.1. Replication Stress

Replication stress is the primary source of genome instability (Aguilera and Garcia-Muse, 2013) and is a major source of human disorders including cancer, growth defects and

neurodegeneration (Fragkos and Naim, 2017). The primary reasons for instability associated with replication stress is debated (Helmrich *et al.*, 2013, Gaillard and Aguilera, 2016), but generally requires transcription (Aguilera, 2002) and is often associated at 'hotspots', including tRNA genes in yeast (Admire *et al.*, 2006). A common feature of replication stress is the pausing or subsequent stalling of the replication fork at actively transcribed genes, creating the conditions that permit the formation of single-stranded DNA, associated DNA breaks and subsequent repair through recombination. The scattering of tRNA genes throughout the yeast genome may well provide a form of protection against instability issues arising from their presence: **Figure 1.3** illustrates the genomic spacing of tRNA genes and their associated transposable elements.

1.4.2. Collisions Between the Replication Fork and Transcriptional Machinery

As the transcriptional and replicative machineries compete for the same template of DNA, it is logical that these two complexes compete and interfere with each other. Collisions between these two machineries are associated with DNA polymerase stalling, replication fork collapse and DNA breaks (Prado and Aguilera, 2005, Gottipati *et al.*, 2008). tRNA genes, due to their highly-transcribed nature, are no exception: they have been shown to interfere with progression of the replication fork resulting in replication stress leading to chromosome breakage and recombination (Admire *et al.*, 2006). Common to tRNA genes are the phenomena of replication fork pausing (or a slowing without collapse): the helicase, Rrm3p, is responsible for the progression of the replication fork following pausing at tRNA genes in yeast and its removal is associated with increased replication fork breakage events (Ivessa *et al.*, 2003). Yeast appears to have evolved a global mechanism to reduce tRNA gene transcription, *via* components associated with the replication stress checkpoint pathway and the regulator Maf1, in response to replication stress caused by tRNA gene transcription (Nguyen *et al.*, 2010).

Cells without the global negative-regulator Maf1 were shown to be sensitive to replicative stress.

The precise mechanism by which transcription/replication fork collision events promote instability is debated (Gaillard and Aguilera, 2016), but may involve barriers on the DNA template including bound proteins such as the transcriptional machinery itself. These barriers may hinder progression of the replication fork, leading to replication fork collapse and repair by the mechanisms described above.

1.4.3. The Role of R-Loops in Promoting Replication Stress

DNA damage caused by collisions between the replication fork and transcriptional machinery have long thought to drive genomic instability. However, the last decade has seen accumulating evidence for the role of RNA/DNA hybrids (R-loops) (Aguilera and Garcia-Muse, 2012) in promoting DNA damage and replication stress. R-loops are a three-stranded RNA/DNA hybrid likely caused by an association between transcribed mRNA and the template strand of DNA (the "thread-back" model), leading to exposure of a single-stranded loop of the non-template strand (**Figure 1.2**). R-loops have been observed to form preferentially at highly-transcribed, GC-rich regions: studies into the genome-wide organisation of R-loops show an association with hotspots of the yeast genome (El Hage *et al.*, 2014), with a strong detection observed at tRNA genes.

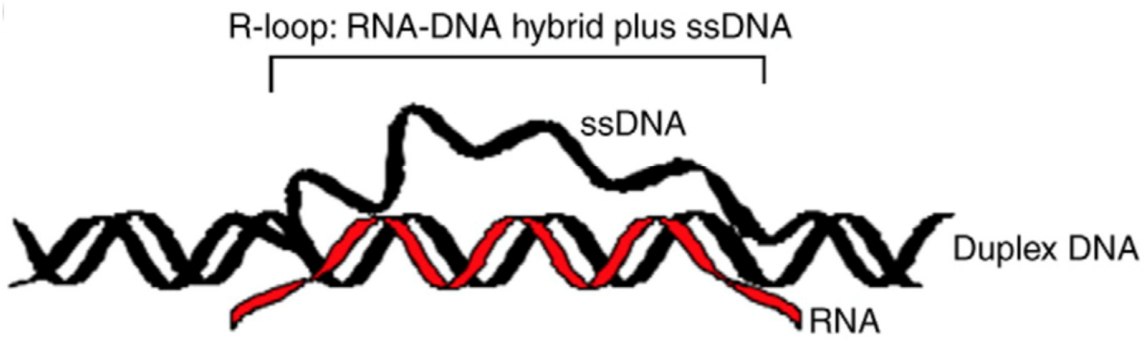


Figure 1.2: Representative diagram of R-loop formation at actively transcribed genes. Following transcription, the nascent RNA transcript is proposed to thread-back and anneal with DNA, leading to the formation of ssDNA. The above image was adapted from Ivancic-Bace *et al.* (2012).

Significant evidence suggests that the presence of R-loops result in DNA breaks leading to genomic instability and directly interfere with progression of the replication fork (Aguilera and Garcia-Muse, 2012, Sollier and Cimprich, 2015, Chang and Stirling, 2017). However, only recently has greater understanding of the mechanistic role of R-loops in promoting replication stress come to light.

If not removed, R-loop formation may cause instability in at least two ways. In humans, exposed ssDNA can be targeted by nucleotide excision repair factors associated with transcription (Sollier *et al.*, 2014) or DNA-modifying enzymes resulting in DNA damage (Ruiz *et al.*, 2011). Alternatively, the R-loop can lead to DNA replication stress and DNA damage by directly impeding the replication fork (Chan *et al.*, 2014, Madireddy *et al.*, 2016). It has recently been suggested that defects in replication fork protection factors stabilise R-loop structures leading to replication-transcription conflicts and subsequent replisome stall events (Chang and Stirling, 2017).

Supporting evidence for the role of R-loops in promoting genome instability include studies into RNA processing enzymes, primarily that of the conserved RNase H enzyme

pair that is responsible for removing the RNA component from the RNA:DNA hybrid. The removal of R-loops by RNase H enzymes is necessary to prevent genomic instability (Lin *et al.*, 2010) and deletion of RNase H in yeast is associated with increased R-loop formation and genome instability (Wahba *et al.*, 2011). Additional evidence suggests that topoisomerase increases genomic stability by preventing R-loop formation and interference between replication and transcription because of negative torsional stress (Tuduri *et al.,* 2009). Loss of topoisomerase I in budding yeast has also been shown to produce R-loops in the rDNA locus resulting in transcriptional blocks (El Hage *et al.*, 2010). Further links between DNA replication and R-loop formation also include a genome-wide study in hyper-recombinant THO-complex yeast mutants that shows how RNase H1 overexpression leads to a reduction in Rrm3 helicase clusters at actively transcribed genes (Gomez-Gonzalez *et al.*, 2011), suggesting a mitigating role for RNase H following replication fork pausing. Additionally, the hyper-recombination phenotype in this mutant background requires transcription during S-phase (Wellinger *et al.*, 2006), providing further evidence for the link between transcription and associated genome instability.

1.4.4. Replication Fork Stalls are Polar in Nature

Although work by Sekedat *et al.* (2010) suggests that replication fork stall events are independent of transcriptional orientation, a greater body of evidence supports the hypothesis that replication fork stall events are influenced by the direction of transcription in eukaryotes. The primary evidence originated from work by Deshpande and Newlon (1996), who clearly show the polar nature of replication fork pausing at tRNA genes in budding yeast. After performing 2D electrophoresis, they show an accumulation of replication intermediates when orientating an origin to fire onto a tRNA gene facing the opposing orientation. In contrast, this pausing was diminished when both machineries face the same orientation. They suggest that the pausing is caused by collisions between the respective machineries, but also suggest an alternative model

where polar replication stalling is caused by an accumulation of positive supercoiling in the DNA template during transcription. In contrast, if the respective transcriptional and replication machineries are unidirectional, these effects may cancel out. Directionality is also important to genes transcribed by RNA pol II in yeast: replication fork stalling and transcription-associated recombination was observed to be far less pronounced when both travelled in the same orientation (Prado and Aguilera, 2005). Finally, it is notable that replication termination sites in the rDNA locus (discussed in further detail in Chapter 3) ensure that the replication fork always moves in the same direction as that of transcription (Rothstein *et al.*, 2000).

1.4.5. tRNA-Associated Retrotransposons Indirectly Contribute to Genomic Instability

Retrotransposons are genetic elements related to retroviruses (Lerat and Capy, 1999) with five families (Ty1 to Ty5) found in *S. cerevisiae*. Retrotransposons, through their preferential integration in regions proximal to tRNA genes (Ji *et al.,* 1993), also play an indirect role in genome instability. These include the retrotransposons themselves, which may form double strand breaks through perturbed (slowed) replication between retrotransposon pairs (Lemoine *et al.,* 2005), or their associated highly-homologous repetitive elements (solo LTRs; n=268) that present hotspots for homologous recombination (Mieczkowski *et al.*, 2006). **Figure 1.3** displays the proximity of known retrotransposons in *S. cerevisiae* with associated tRNA genes.

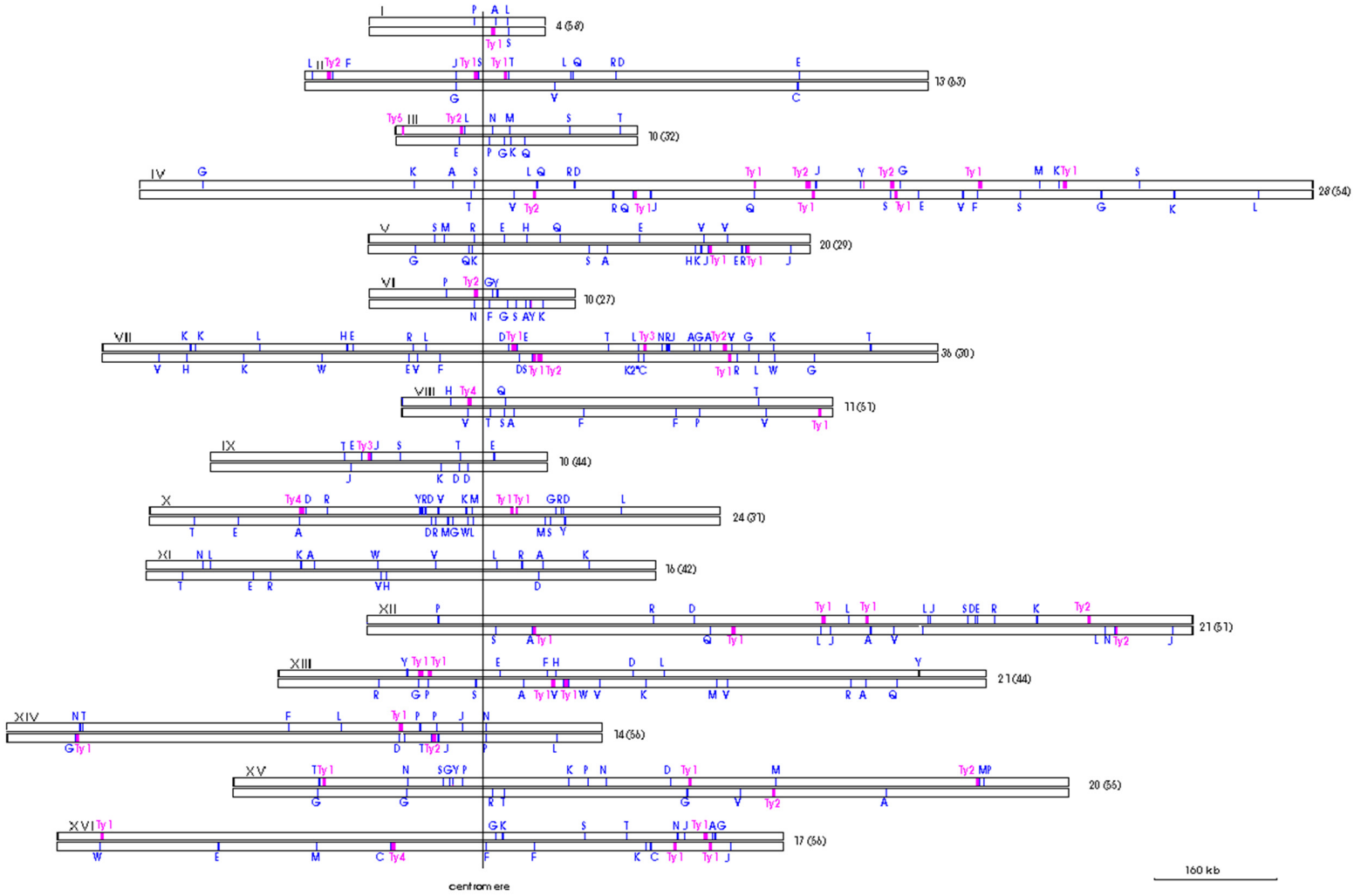


Figure 1.3: The scattering of tRNA genes and associated retrotransposons throughout the sixteen yeast chromosomes. The figure shows the proximity of tRNA genes (blue) with their associated retrotransposons (pink). Each tRNA gene is represented by a single-letter code corresponding to each amino acid, with the retrotransposon classes (*Ty1* to *Ty4*) indicated. The line through the above image indicates the position of the centromere. The above figure was adapted from Hani and Feldmann (1998).

1.5. Overview of Chromosome Biology

To be considered a true chromosome, a synthetic neochromosome must consist of a series of individual features that are critical for function *in vivo*. A brief overview of the fundamentals of chromosome biology is presented here, and includes a description of the centromere, telomeres and chromosome replication. Additionally, as the tRNA neochromosome was initially constructed as a circular 'megaplasmid' prior to linearisation using the telomerator system (described in Chapter 5), this section also presents a discussion on the fundamental behaviour of ring chromosomes in eukaryotes.

1.5.1. Fundamental Features of a Eukaryotic Chromosome

Eukaryotic genomes consist of a series of structures called chromosomes, with the genome of *S. cerevisiae* consisting of sixteen chromosomes. To ensure proper segregation during mitosis, each chromosome contains a centromere where microtubules attach, *via* the kinetochore, to ensure equal segregation during mitosis (Fitzgerald-Hayes *et al*., 1982). The structure of the centromere domain often differs between even closely related eukaryotes: in contrast to *Saccharomyces pombe* that contains large centromere sequences of several kilobases and many repetitive elements, *S. cerevisiae* contains a short, 125 bp AT-rich consensus region with no repetitive DNA (Clarke, 1990).

A second defining feature of eukaryotic chromosomes are the telomeres. Telomeres in *S. cerevisiae* consists of approximately 300 bp tandem TG-rich repeats (Shampay *et al.*, 1984) that protect chromosome ends from progressive shortening (the end-replication problem) and chromosome fusion. *S. cerevisiae* counteracts the progressive shortening of telomers by expressing a specialised reverse-transcriptase, telomerase, that extends telomere arms (de Lange, 2009). Telomeres also assist in the maintenance of the three-dimensional structure of the yeast genome by tethering chromosomes to the nuclear

envelope (Kupiec, 2014). This tethering may also play a role in the silencing of genes proximal to telomeres through the telomeric position effect (TPE) (Sandell and Zakian, 1992). Lastly, synthetic telomere seed sequences may also be added to synthetic circular chromosomes using the telomerator system (Mitchell and Boeke, 2014) and is discussed in further detail in Chapter 5.

Replication of eukaryotic DNA is a highly-complex process that requires the coordination of multiple processes. DNA replication in *S. cerevisiae* is initiated by a series of scattered origins of replication (ARS elements), defined by their ability to independently support plasmid replication in yeast (Stinchcomb *et al.*, 1979). Each origin consists of a 11 to 17 bp thymidine-rich ARS Consensus Sequence (ACS) (Deshpande and Newlon, 1992) that are bound by the Origin Recognition Complex (ORC) throughout the cell cycle. The ORC is responsible for the initiation of DNA synthesis in S-phase through formation of the multi-protein prereplicative complex (Aparicio *et al.*, 1997). The MCM2-7 (Mini Chromosome Maintenance protein) forms the core of the replicative helicase that itself forms the body of a replication fork that progresses along DNA (Vijayraghavan and Schwacha, 2012).

An unusual feature of eukaryotic chromosomes are replication termination sites located in the rDNA locus. Although there exists an extensive body of work on genome replication, very little is understood about the nature of replication termination. In *S. cerevisiae*, Fob1 is the primary replication fork blocking (RFB) protein, and acts by blocking replication in the rDNA locus by binding to *Ter* sites (Huang and Maraia, 2001). They appear to be responsible for preventing collisions between the replicative and transcriptional machinery in the highly-transcribed rDNA locus (Takeuchi *et al.*, 2003). Transcriptional directionality relative to the replication fork also appears to be critical: in all observed cases, *Ter* sites ensure that the replication fork travels in the same direction as transcription (Rothstein *et al.*, 2000).

1.5.2. The Unstable Behaviour of Ring Chromosomes in Eukaryotes

By their very nature, circular chromosomes present inherent topological challenges for eukaryotes. Yeast, as do all eukaryotes, have evolved with linear chromosomes protected by telomeres, and generally lack the mechanisms evolved by bacteria to resolve circular genomes. For example, yeast lack the bacterial ATP-dependent gyrase that is responsible for the unwinding of circular DNA supercoiling (Dorman and Dorman, 2016). Additionally, bacteria have evolved responses to resolve recombination-dependent catenane formation between circular chromosomes: circular monomeric bacterial chromosomes are known to dimerise through homologous recombination and are resolved back to monomers through Xer site-specific recombination at the *dif* site (Barre *et al.*, 2001).

Ring chromosomes in higher eukaryotes are associated with diseases, including cancer (Gebhart, 2008, Gisselsson *et al.*, 2000) and autosomal "ring syndromes" in humans (Yip, 2015, Gisselsson, 2002). Phenotypes of ring syndromes vary, but may include epilepsy, intellectual disability, dwarfism and Turner syndrome (Berkovitz *et al.*, 1983). Ring chromosomes in eukaryotes are generally unstable, and repeatedly fuse and break during cell division leading to the formation of highly-variable and genetically distinct daughter cells (the breakage-fusion-bridge cycle). In such cases, sister chromatid exchange is the primary source of instability, leading to the formation of dicentric chromosomes (a chromosome with two centromeres), or catenane (interlocked ring) formation (Yip, 2015). For example, the presence of mosaicism (or dynamic mosaicism) has been described for ring chromosome 4, characterised by a high degree of dynamic variability and structural instability following sister chromatid exchange (McDermott *et al.*, 1977).

The unstable behaviour of ring chromosomes in eukaryotes was described almost 80 years ago by the seminal studies in maize of McClintock (1938). In contrast to linear chromosomes, circular chromosomes may undergo mitosis in three ways based on the

number of sister chromatid exchange events (Gisselsson, 2002, Yip, 2015). If no sister chromatid exchange takes place, circular chromosomes will segregate normally during mitosis. However, if two sister chromatid exchanges occur, an interlocked ring (catenane) will be formed. Alternatively, a dicentric ring will be formed in the event of a single chromatid exchange. It should be noted that DNA catenation is not solely restricted to circular chromosomes, and is also a natural consequence of DNA replication for linear chromosomes (Murray and Szostak, 1985, Uemura *et al.*, 1987, Charbin *et al.*, 2014).

Anaphase bridges may form following catenane or dicentric chromosome formation, leading to DNA tearing and an uneven exchange of genetic material during cell division. This instability may also be cyclical if the linear broken ends are repaired through fusion (McClintock, 1938, Gisselsson et al., 2000), resulting in the classic breakage-fusion-bridge cycle and genetic variability of ring chromosomes in eukaryotes. A proposed model for circular tRNA neochromosome behaviour may be found in Chapter 6.

Finally, Xie *et al.* (2017) also studied the behaviour of a ring variant of synthetic chromosome V. For both native and synthetic ring derivatives, they observe a restoration in the expression of subtelomeric genes indicating a loss of telomere silencing. While they observe no change in sporulation frequency, they observe a reduction in the viability of spores suggesting a mis-segragation during meiosis, possibly caused by the formation of dicentric or catenane chromosomes. It should also be noted that Xie *et al.* (2017) did not characterise the structural stability of the ring SynV variants during mitosis, and rather investigated the re-formation of linear SynV variants following ~60 generations of strains housing the ring SynV variant in liquid YPD media.

1.6. Motivation and Objectives of this Study

In their paper, Dymond *et al.* (2011) introduced the aims for constructing a synthetic yeast genome. The overall basis of the Sc2.0 consortium is formed around three central design principles:

- *First, it should result in a (near) wild-type phenotype and fitness*;
- *Second, it should lack destabilizing elements such as tRNA genes or transposons; and*
- *Third, it should have genetic flexibility to facilitate future studies*.

To meet the objectives of the second design principle, Dymond *et al.* (2011) describe the intention of the Sc2.0 consortium to remove and relocate all 275 tRNA genes onto a dedicated extrachromosomal array (*i.e.* a neochromosome) to reduce any instability resulting from their presence in the genome.

It was the objective of this doctoral project to design, construct and characterise a tRNA neochromosome housing all 275 tRNA genes that will eventually complement their absence in a synthetic yeast strain. This neochromosome is expected to form an essential component of the overall Sc2.0 project and additionally provide a resource, in the form of tRNA arrays, to consortium members in different stages of synthetic chromosome construction.

As summarised in Section 1.4, tRNA genes are known hotspots for genomic instability. The removal of tRNA genes from the synthetic yeast genome and their collation onto a dedicated chromosome will therefore theoretically result in a genome more robust to structural rearrangement and instability. Secondary objectives (not part of the scope of this doctoral study) are to provide a chassis to study fundamental aspects of tRNA biology through the Dre-*rox* recombinase system and a future means to understand the fundamentals of chromosome biology.

At the time of writing, the tRNA neochromosome likely represents the first attempt at constructing a true, fully-synthetic, designer neochromosome *de novo*. Due to its novel nature, the technologies and approaches described in this thesis are previously untested. A further objective of this doctoral study was to develop and test enabling technologies to provide researchers with the means to construct neochromosomes in future.

Synthetic biology may be described as an inter-disciplinary amalgamation of biology and engineering. As this project essentially involves engineering at the molecular level, an engineering approach formed a central theme throughout much of this thesis. Common to synthetic biology and engineering is the ubiquitous "Design-Build-Test" cycle (Liu *et al.*, 2015). The body of this thesis is therefore split into three major chapters:

1. **Design** (Chapter 3: Designing the tRNA Neochromosome): This chapter forms an important component of the neochromosome project and is critical to ensure its stability and overall functionality *in vivo*. Design principles are strongly influenced by engineering concepts and are intended to rationalise the overall approach towards design, maximise neochromosome stability and ensure components function as expected.
2. **Build** (Chapter 4: Constructing the tRNA Neochromosome): This chapter describes the development of new technologies to construct the neochromosome and reliably verify integration of each individual component. Additional objectives include the development of technologies to enable future neochromosome construction.
3. **Test** (Chapter 5: Characterising the tRNA Neochromosome): This chapter describes the functional verification of individual neochromosome components and verification that the neochromosome is indeed an intact, physical object. This work includes the characterisation of tRNA genes on an individual level, origins of

replication and replication termination elements. Additional work includes the physical manipulation of the neochromosome structure, including linearisation and chemical transfer from one cell to another. Finally, deep sequencing is critical to determine the overall fidelity of the neochromosome sequence.

Chapter 2: Materials and Methods

2.1. Chemicals, Enzymes and Oligonucleotides

Unless otherwise stated, all chemicals were obtained from Sigma-Aldrich (Dorset, UK). Enzymes (restriction enzymes, DNA polymerase *etc.*) were obtained from New England Biolabs (Hitchin, UK) unless otherwise stated. All oligonucleotides were synthesised by Integrated DNA Technologies (Leuven, Belgium).

2.2. *De novo* DNA Synthesis

DNA sequences were synthesised *de novo* by external vendors as follows. Neochromosome DNA, in the form of tRNA arrays, was synthesised by Wuxi Qinglan Biotech Co. Ltd (Yixing City, China), except for the Chr8, Chr9 and Chr16 tRNA arrays that were synthesised by Thermo Fisher Scientific (Renfrew, UK). *Ter* sites were synthesised by Twist Biosciences (San Francisco, USA) and short stretches of DNA ('gBlocks') were synthesised by Integrated DNA Technologies (Leuven, Belgium).

2.3. Media

LB media (broth and plates), YPD media (broth and plates) and SC dropout media (plates and dropout stocks) were prepared by Mrs. Aileen Greig (Edinburgh, UK).

2.3.1. Luria-Bertani (LB) Media

LB media was comprised of the following ingredients:

- 10 g/L tryptone (BD: Becton, Dickinson and Company; Oxford, UK)
- 5 g/L yeast extract (BD; Oxford, UK)

- 10 g/L sodium chloride
- 15 g/L agar (if plates required)

The above was prepared up to appropriate volumes with ddH_2O and autoclaved at 121°C for 15 minutes. For ampicillin selection in bacteria, 100 mg/mL stocks of ampicillin disodium salt were added for a final concentration of 100 µg/mL once media had cooled to below 50°C. For kanamycin selection in bacteria, 50 mg/mL stocks of kanamycin sulphate were added for a final concentration of 50 µg/mL once media had cooled to below 50°C.

2.3.2. Yeast Extract, Peptone, Dextrose (YPD) Media

YPD media was comprised of the following ingredients:

- 10 g/L yeast extract (BD; Oxford, UK)
- 20 g/L peptone (BD; Oxford, UK)
- 20 g/L dextrose (Merck Millipore; Feltham UK)
- 15 g/L agar (if plates required)

The above was prepared up to appropriate volumes with ddH_2O and autoclaved at 121°C for 15 minutes. For *KanMX* selection in yeast, 50 mg/mL stocks of G418 (Geneticin; Alpha Diagnostics International) were added for a final concentration of 200 µg/mL once media had cooled to below 50°C.

2.3.3. Synthetic Complete Media

Synthetic complete (SC) media was initially prepared from 2X SC amino acid drop-out stock solutions. **Table 2.1** describes the ingredients used to prepare SC-8 drop-out amino acid mix.

Table 2.1: Amino acids used to prepare SC-8 drop-out amino acid mix.

Supplement	Amount
L-Alanine	10 g
L-Arginine-HCl	10 g
L-Asparagine	10 g
L-Aspartic acid	10 g
L-Glutamine	10 g
L-Glutamic acid	10 g
Glycine	10 g
Myo-inostiol	10 g
L-Isoleucine	10 g
L-Phenylalanine	10 g
L-Proline	10 g
L-Serine	10 g
L-Threonine	10 g
L-Tyrosine	10 g
L-Valine	10 g
Total weight	*150 g*

To prepare -6 dropout powder, 8.4 g of L-Methionine and 8.4 g of L-Cysteine was added to 126g of the above SC-8 dropout mix. For SC-Ura-Trp dropout powder, 1.1 g of Adenine, 4.4 g of L-Histidine, 8.8 g of L-Leucine and 4.4 g of L-Lysine was added to 74.8 g of -6 dropout powder. For SC-His-Leu dropout powder, 0.7 g of Adenine, 2.8 g of L-Tryptophan, 2.8 g of Uracil and 2.8 g of L-Lysine was added to 47.6 g of -6 dropout powder.

2X dropout stocks were comprised of the following ingredients:

- 3.4 g/L Yeast nitrogen base (without amino acids and ammonium sulphate) (BD; Oxford, UK)

- 10 g/L Ammonium sulphate (VWR; Leicestershire, UK)

For 2X -His-Leu dropout stock, 4 g/L -His-Leu drop out mix was added. For 2X -Ura-Trp dropout stock, 4 g/L -Ura-Trp drop out mix was added. For 2X SC-8 dropout stock, 4 g/L SC-8 dropout mix (lacking Ura, Trp, Leu, His, Lys, Ade, Cys and Met) was added.

To prepare 500 mL of Synthetic Complete media from 2X dropout stocks, 50 mL of sterile 20% glucose or 50 mL of 20% galactose (filter-sterilised; 0.22 µM) was added to 200 mL of 2X dropout stock, with additional amino acids added where required (**Table 2.2**).

Table 2.2: Amino acid supplement solutions used to prepare Synthetic Complete dropout media. The individual components were included or omitted where required.

Extra Amino acid solutions	Volume needed for 500mL
100mM Leucine	10 mL
40mM Tryptophan	5 mL
20mM Uracil	5 mL
16mM Adenine	5 mL
100mM Lysine	5 mL
100mm Histidine	1.5 mL
50 mM Methionine	6 mL

Volumes were then adjusted to a final volume of 500 mL with sterile ddH_2O. If plates were required, 250 mL of molten 4% agar was added to the above and carefully mixed before pouring. For 5-Fluoroorotic Acid Monohydrate (5-FOA) counter-selection against the *URA3* marker in yeast, 100 mg/mL stocks of 5-FOA (Formedium) in absolute dimethyl sulphoxide were added for a final concentration of 1 mg/mL once media had cooled to below 50^{o}C.

2.4. *S. cerevisiae* Strains Used in this Study

A full list of yeast strain genotypes used in this study may be found in Appendix V, **Table IV**.

Table 2.3: Yeast strains used in this study. The following two *S. cerevisiae* strains were used to construct the tRNA neochromosome in parallel. BY4741 was used for all other purposes, including yeast assembly through *in vivo* homologous recombination.

Strain ID	Name	Genotype	Source
RWy007	BY4741	MAT**a** *leu2*Δ0 *met15*Δ0 *ura3*Δ0 *his3*Δ1	Laboratory collection, Cai Lab, UK
RWy121	SynIII/VI/IXR	MAT*α* *leu2*Δ0 *lys2*Δ0 *MET15* *ura3*Δ0 *his3*Δ1 synIII *ho*::Syn.*SUP61*::*ura3* synVI SYN-WT.*PRE4* IXL-synIXR	Jef Boeke, NYU School of Medicine, USA

2.5. *E. coli* Strain Used in this Study

Escherichia coli DH5α (genotype: F^- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rk^-, mk^+) *pho*A *sup*E44 λ^- *thi*$^-$1 *gyr*A96 *rel*A1) was procured from Invitrogen and used to prepare competent cells.

2.6. Manipulation of Strains

Manipulation of bacteria and yeast strains was undertaken using standard microbiology techniques under sterile conditions. For long-term storage, all strains were stored at -80^{o}C in liquid culture media to a final concentration of 15% glycerol.

2.7. Molecular Biology Techniques

2.7.1. Bacteria Competent Cell Preparation

Competent bacteria were prepared using the rubidium chloride method according to Green and Rogers (2013), with modifications noted as follows.

A pre-culture of *E. coli* was prepared by inoculating 5 mL of LB and incubated with rotation in a Cell Culture Roller Drum (Eppendorf) at 37^{o}C overnight. Overnight cultures were then re-inoculated into 250 mL of LB medium containing 20 mM $MgSO_4$ and incubated with shaking (140 rpm; Infors HT Multitron Standard) at 37^{o}C until OD_{600nm} reached between 0.4 to 0.6. Cells were then pelleted at 4,500 g for 5 minutes at 4^{o}C (Heraeus Multifuge X3 FR) and re-suspended in 0.4 volumes of TFB1 (30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl and 15% glycerol; adjusted to pH 5.8 and filter-sterilised (0.22 µM)). Cells were incubated on ice for 5 minutes, centrifuged at 4,500 g for 5 minutes and re-suspended in 1/25th volume of ice-cold TBF2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl and 15% glycerol; adjusted to pH 6.5 and filter-sterilised (0.22 µM)). Cells were then incubated on ice for 15 to 60 minutes, aliquoted in 50 µL volumes, flash-frozen on dry ice and stored at -80^{o}C for future transformation.

2.7.2. Bacterial Transformation

Bacterial transformation was performed according to Mandel and Higa (1970), with modifications noted as follows.

Competent cells (50 µL aliquots) were initially thawed on ice before the addition of 5 µL of DNA and briefly mixed. Tubes were incubated on ice for 20 minutes before being subjected to heatshock at 42^{o}C for 45 seconds and then incubated on ice for two minutes. 0.5 mL of LB media was added to each tube and incubated with shaking at 900 rpm in a Thermomixer (Eppendorf) at 37^{o}C for one hour. Cells were then centrifuged for 10 seconds at 15,000 g with the resulting pellet re-suspended in 100 µL of the remaining supernatant. Transformations were serially-diluted if necessary, plated onto LB + Amp or LB + Kan plates and incubated overnight at 37^{o}C.

2.7.3. Bacteria 'Miniprep'

Plasmid extraction from *E. coli* cells was undertaken by inoculating bacteria isolates into 5 mL of LB + Amp (100 µg/mL) and incubating with rotation in a Cell Culture Roller Drum (Eppendorf) at 37^{o}C overnight. Plasmids were extracted using the QIAprep Spin Miniprep Kit (Qiagen) according to manufacturer's instructions and quantified using the NanoDrop Spectrophotometer (Thermo Scientific).

2.7.4. Yeast Transformation

Yeast transformation was performed according to Gietz and Woods (2002), with modifications noted as follows. All solutions were sterilised by filtration (0.22 µM) or autoclaved at 121^{o}C for 15 minutes.

A pre-culture of yeast was prepared by inoculating single isolates into 10 mL of liquid media and incubated with rotation in a Cell Culture Roller Drum (Eppendorf) at 30°C overnight. Overnight cultures were then re-inoculated into fresh media to a target OD_{600nm} of 0.1 in 20 mL of liquid media and incubated with rotation until OD_{600nm} reached between 0.5 and 1. Cells were then centrifuged at 2,103 g (Megafuge X3 FR, Heraeus Instruments) for 5 minutes, washed with 10 mL sterile ddH_2O and centrifuged again. Cell pellets were washed with 10 mL of 0.1 M LiOAc, centrifuged at 2,103 g for 5 minutes, with the remaining cell suspension at the bottom used for transformation in a 1.5 mL centrifuge tube.

Yeast transformation was undertaken by adding 5 to 10 µL of transforming DNA*, 36 µL of 1M LiOAc, 19 µL ddH_2O, 25 µL 10 mg/mL herring sperm carrier DNA (Promega; heat-denatured by incubating at 100°C for 10 minutes and cooled on ice before use) to 240 µL of 44% PEG 3,350 and mixed by vortexing. 50 µL of competent yeast cells were then added, mixed by tube inversion, and incubated at 30°C for 30 minutes. 36 µL of DMSO was added to each transformation, quickly mixed, and incubated at 42°C for 15 minutes. Transformations were then centrifuged at 1,300 g for 30 seconds, re-suspended in 400 µL of 5 mM $CaCl_2$ and incubated at room temperature for 10 minutes. Serial dilutions were prepared if necessary, plated onto the required selective media, and incubated for 2-3 days at 30°C.

*DNA concentrations for yeast transformation are variable. For plasmid-based transformations, 20 ng (or less) to 100 ng were used. For integrative transformations, DNA concentrations measured on the Nanodrop spectrophotometer (Thermo Scientific) are unreliable due to the gel purification method used (Qiagen).

*2.7.5. Gene Knock-Out (*KanMX *KO)*

Gene knock-out was undertaken by generating specialised cassettes based on the *KanMX* gene conferring resistance to G418. The *KanMX* gene was first amplified by PCR and flanked with *loxP* sites for subsequent gene removal if required. The PCR product was then TOPO cloned into a TOPO vector, with intervening DNA sequences surrounding the *KanMX* gene amplified using primers containing 40 bp homology arms to the gene of interest. The resulting PCR products were then transformed into yeast. To facilitate recovery when selecting for G418-resistant isolates, transformations were plated onto YPD agar, incubated for one day at 30^{o}C, and replica-plated onto YPD agar containing G418 (200 µg/mL). Successful gene knock-out was then verified using colony PCR with primers designed to flank the junction regions of integration.

2.7.6. Yeast Spheroplast Transformation

Yeast spheroplast transformation was performed according to Burgers and Percival (1987), with modifications noted as follows. All solutions were sterilised by filtration (0.22 µM) or autoclaved at 121^{o}C for 15 minutes.

A pre-culture of yeast was prepared by inoculating single isolates into 10 mL of media and incubated with rotation in a Cell Culture Roller Drum (Eppendorf) at 30^{o}C overnight. Overnight cultures were re-inoculated into 20 mL of fresh media to a target OD_{600nm} of 0.3 and incubated with rotation at 30^{o}C till OD_{600nm} reached 1.0. Cultures were then centrifuged for 5 minutes at 2,103 g (Megafuge X3 FR, Heraeus Instruments) with cell pellets washed with 20 mL sterile ddH_2O, centrifuged again, and washed with 20 mL of sterile 1 M sorbitol. After centrifuging again, cell pellets were re-suspended in 20 mL of SCEM (1 M sorbitol, 0.1 M sodium citrate (pH 5.8) and 10 mM EDTA). To spheroplast cells, 40 µL of 2-mercaptoethanol and 1,000 units of lyticase (Sigma) were added to each tube,

gently mixed, and incubated at 30°C for 20 minutes. The resulting spheroplasts were centrifuged (500 g for 10 minutes), gently washed once with 20 mL of 1 M sorbitol and gently washed again with 20 mL of STC (1 M sorbitol, 10 mM Tris-HCl (pH 7.5) and 10 mM $CaCl_2$). The resulting pellet was re-suspended in 2 mL of STC buffer containing 30 µL of herring sperm DNA (Promega; previously incubated at 100°C for 10 minutes).

To maximise transformation efficiency, increased volumes were used during transformation. 12 µL of neochromosome DNA was added to 300 µL of spheroplasts and incubated at room temperature for 10 minutes before the addition of sterile PEG (10 mM Tris-HCl (pH 7.5), 10 mM $CaCl_2$ and 20% PEG 8,000). Spheroplasts were incubated for 10 minutes, centrifuged (500 g for 10 minutes) and re-suspended in 150 µL of SOS media (1 M sorbitol, 6.5 mM $CaCl_2$, 0.25% yeast extract, 0.5% bactopeptone, supplemented with 80 mg/L tryptophan; filter-sterilised) and incubated at 30°C for 30 minutes. Spheroplasts were then added to 10 mL of molten SC-Ura TOP agar (SC-Ura media containing 1 M sorbitol and 2.5% agar; held at 45°C), gently mixed, and poured directly onto petri dishes containing 10 mL of SC-Ura SORB (SC-Ura media containing 0.9 M sorbitol, 2% agar and 3% glucose). The plates were then incubated at 30°C for four days before the appearance of sufficiently large colonies.

2.7.7. Yeast Plasmid Extraction

Plasmid extraction and purification from yeast was undertaken according to the following internal laboratory protocol ("Roy's Magic Yeast Miniprep").

A single yeast isolate was inoculated into 10 mL of selective media and incubated with rotation in a Cell Culture Roller Drum (Eppendorf) at 30°C overnight. 1.5 mL of cells were then centrifuged at 600 g for 3 minutes in 1.5 mL centrifuge tubes and re-suspended in 200 µL of solution 1 (1 M sorbitol, 10 mM Tris-HCl (pH 7.5), 10 mM EDTA (pH 8.0)) with 3

µL of Zymolyase solution (1% 2-mercaptoethanol, Zymolyase (Cambridge Bioscience, Cambridge; 5 units per µL in sterile glycerol (60%)) added to each reaction. Cells were then spheroplasted by incubating at 37°C for 1 hour. 250 µL of buffer P2 (200 mM NaOH and 1% SDS) was then added to lyse cells, mixed well, and neutralized with the addition of 350 µL of Buffer N3 (Qiagen). The resulting lysate was centrifuged at 17,000g for 3 minutes, with the supernatant centrifuged again in fresh 1.5 mL centrifuge tubes to remove residual protein debris. The supernatant was transferred to EZ-10 DNA spin columns (NBS Biologicals), centrifuged at 17,000g for 1 minute, and washed by adding 700 µL of buffer PE (10 mM Tris-HCl (pH 7.5) in 80% ethanol) and centrifuged again at 17,000g for 1 minute. To remove any traces of ethanol, spin columns were centrifuged a final time prior to elution with 50 µL of 1X EB buffer (10 mM Tris-HCl (pH 7.5)). The DNA solution was then vacuum-concentrated (Eppendorf Concentrator Plus, Eppendorf) at 45°C for 30 minutes and re-suspended in 12 µL of sterile ddH_2O. 5 µL of the resulting DNA was then transformed into chemically competent *E. coli*.

2.7.8. Restriction Digests

Restriction digests were performed on purified plasmid DNA or gel-purified PCR products. Appropriate volumes of DNA sample were used in 25 µL reactions, supplemented with 2.5 µL of the corresponding buffer (New England Biolabs) and H_2O to the target volume. 1 µL of restriction enzyme (New England Biolabs) was used per sample, and incubated at the appropriate temperature for 1 hour. Restriction digests were then analysed using agarose gel electrophoresis, with DNA fragments excised and purified where necessary.

For restriction digest DNA fingerprinting, plasmids were digested using specific restriction enzymes to produce a distinct pattern on an agarose gel. Constructs were deemed to be positive or negative based on the resulting gel pattern.

2.7.9. DNA Ligation

Ligation of DNA fragments subjected to restriction digest was performed with T4 DNA ligase according to manufacturer's instructions (NEB: New England Biolabs). Reaction volumes of 20 µL were prepared by combining vector and insert DNA (using a molar ratio of between 3:1 and 10:1 of insert to vector) before adding 2 µL of 10X T4 DNA Ligase Buffer (NEB) and 1 µL of T4 DNA Ligase (NEB). Reaction volumes were then brought up to 20 µL, mixed thoroughly, and incubated at room temperature for two hours. 5 µL of the resulting ligation was then transformed into competent bacterial cells.

2.7.10. Polymerase Chain Reaction (PCR)

PCR reaction conditions were generally optimised for each reaction according to manufacturer's instructions. All PCR reactions were performed in a pre-heated ProFlex PCR Thermal Cycler (Thermo Fisher Scientific) with 32-well or 96-well blocks using either individual 0.2 mL PCR tubes (Applied Biosystems) or 96-well PCR plates (Brand GMBH).

For standard screening conditions, such as that of colony PCR, GoTaq G2 Green Master Mix (Promega) was used. Yeast colony PCR screening was performed according to the internal 'Neta's Nifty Yeast Colony PCR Protocol' (Neta Agmon) by incubating yeast isolates in 50 µL of 20 mM NaOH for 10 minutes, with 1 µL of yeast lysate used as template DNA for each reaction. **Table 2.4** describes the PCR reaction conditions and **Table 2.5** describes the thermal cycler program settings for GoTaq Green.

Table 2.4. PCR reaction conditions for GoTaq Green.

Component	Volume (µL)
GoTaq® Green Master Mix, 2X	6.25
Upstream primer (10 µM)	0.5
Downstream primer (10 µM)	0.5
DNA template	0.5 to 1.5
Nuclease-Free Water	4.75

Table 2.5. Thermal cycler program settings for GoTaq Green.

	Initial Denaturation	Amplification			Final Elongation	Cooling
		Denaturation	Annealing	Elongation		
Cycles	1	30			1	1
Temperature	95°C	95°C	50°C	72°C	72°C	8°C
Time	2 min	30 s	1 min 15 s	1 min 45 s	5 min	hold

For PCR reactions requiring high-fidelity, such as that of amplifying a DNA target of interest, Phusion High-Fidelity DNA Polymerase (New England Biolabs) was used. GC buffer and 3% DMSO were used to increase the efficiency of reactions containing primers with 'tails' (such as Gibson Assembly). **Table 2.6** describes the PCR reaction conditions and **Table 2.7** describes the thermal cycler program settings for Phusion polymerase.

Table 2.6. PCR reaction conditions for Phusion polymerase.

Component	Volume (µL)
Nuclease-Free Water	To 25
5X Phusion GC buffer	5
dNTP mix (10 mM)	0.5
Upstream primer (10 µM)	1.2
Downstream primer (10 µM)	1.2
Phusion DNA Polymerase	0.25
DNA template	1
DMSO (3%)	0.75

Table 2.7. Thermal cycler program settings for Phusion polymerase.

	Initial Denaturation	Amplification			Final Elongation	Cooling
		Denaturation	Annealing	Elongation		
Cycles	1	30			1	1
Temperature	98°C	98°C	*°C	72°C	72°C	8°C
Time	30 s	10 s	30 s	* s	8 min	hold

*Annealing temperature and elongation time were optimized for each reaction

2.7.11. Agarose Gel Electrophoresis

Agarose gels were prepared by adding low-electroendosmosis (EEO) agarose powder (Sigma) to 1X Tris-acetate-EDTA buffer (TAE; 40 mM Tris base, 20 mM acetic acid, 1 mM EDTA). Depending on electrophoresis conditions, appropriate quantities of agarose were added for a final concentration of between 1% to 2%. The agarose was melted by heating in a microwave with regular mixing and cooled to approximately 55°C prior to the addition of a 1 in 10,000 dilution of SYBR safe (Life Technologies). The gel was left to solidify in a gel casting apparatus with a gel comb (Bio-Rad). Once solidified, the gel was placed into a gel tank (Bio-Rad) containing appropriate volumes of 1X TAE buffer.

Half the recommended volume of 6X Purple loading dye (New England Biolabs) was added to each DNA sample prior to loading (lower volumes of loading dye were observed to reduce UV shadow). A 2-log DNA ladder (**Figure 2.1**; New England Biolabs) was used to quantify DNA fragment size according to manufacturer's instructions. Agarose gels were then left to run with 110 volts of electric current until DNA fragments were sufficiently separated. Agarose gels were visualised under UV light (behind a protective XcitaBlue™ UV filter) on a Gel Doc XR+ imaging system (Bio-Rad).

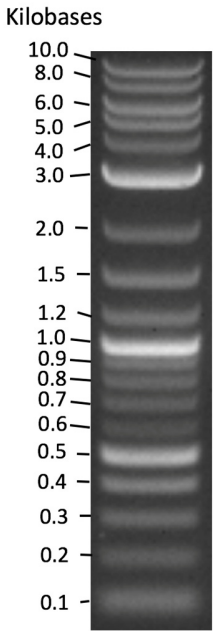


Figure 2.1: Representative 2-log ladder used to quantify DNA fragment size. The 2-log ladder was used on all agarose gel images throughout this thesis. Sizes of each band in kilobases are indicated.

2.7.12. Gel Extraction and Purification of DNA

Gel extraction was undertaken by visualising DNA fragments of interest on a Gel Doc XR+ imaging system (Bio-Rad) and excising the gel fragment using GelX Gel Excision pipette tips (Scientific Laboratory Supplies). Gel purification was then performed by using the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions

2.7.13. TOPO Blunt-End Cloning

TOPO blunt-end cloning of gel-purified PCR products was performed using the Zero Blunt TOPO PCR Cloning Kit (Life Technologies) according to manufacturer's instructions.

2.7.14. Gibson Assembly

DNA fragments with 40 bp overlapping regions were assembled together according to the method described by Gibson *et al.* (2009) ('Gibson Assembly') with differences noted as follows. A schematic overview of Gibson Assembly may be found in **Figure 2.2**.

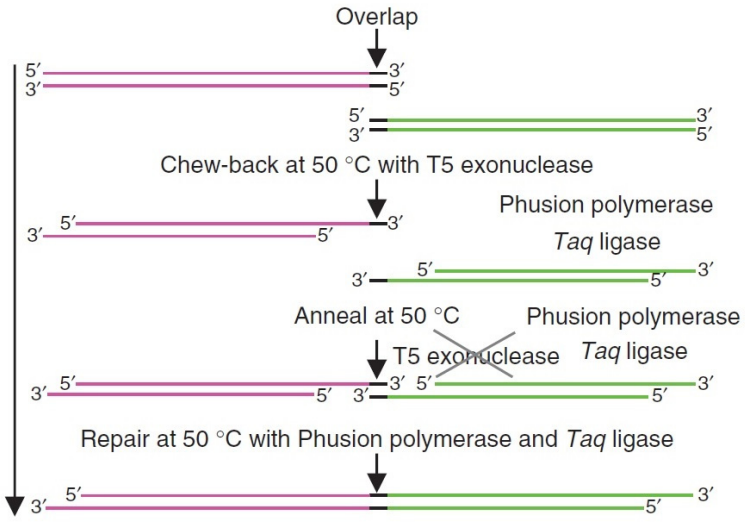


Figure 2.2: Schematic overview of Gibson Assembly. Gibson Assembly is a single-step isothermal reaction designed to assemble DNA fragments with an overlapping homology using three enzymes (T5 exonuclease, a polymerase and a ligase). In the above figure, two DNA fragments (magenta and green) share 40 bp regions of homology (black). T5 exonuclease initially excises the 5' region of each linear DNA fragment and allow complementary annealing between the 40 bp overlapping regions. Phusion polymerase then fills in the gaps before *Taq* ligase seals nicked DNA. The black cross on the above figure indicates the heat-labile nature of T5 exonuclease, which is inactivated during the 50°C incubation. The above image was adapted from Gibson et al. (2009).

Gibson Assembly master-mix was prepared by combining 40 µL of 5X ISO buffer (25% PEG-8000, 500 mM Tris.HCl (pH 7.5), 50 mM $MgCl_2$, 50 mM DTT, 1 mM each of dNTPs, and 5 mM NAD), 1.6 µL of T5 exonuclease (NEB; New England Biolabs), 2.5 µL Phusion polymerase (NEB), 20 µL of Taq ligase (NEB; 40 U/µL) and 85.9 µL of nuclease-free H_2O. The master mix was then aliquoted in 15 µL volumes into 0.2 mL PCR tubes.

To generate DNA fragments with the required regions of homology, primers were designed to contain a 40 bp overlapping 'tail' so that PCR amplification products would overlap to regions of interest. Vector DNA was also first linearized using restriction digest. DNA products were then gel-purified and thoroughly mixed with a 15 µL aliquot of the Gibson assembly master mix (previously thawed on ice) to a final volume of 20 µL. Generally, a 3:1 ratio of insert to vector was used. The Gibson Assembly reaction was then undertaken by incubating at 50^{o}C for 45 minutes on a pre-heated ProFlex PCR Thermal Cycler (Thermo Fisher Scientific). 5 µL of the resulting reaction was transformed into 50 µL of chemically competent *E. coli*. Resulting bacteria colonies were verified using colony PCR and/or restriction digest verification with junction regions screened using Sanger sequencing.

2.7.15. DNA Assembly Using In vivo Homologous Recombination

If Gibson Assembly failed to produce the expected construct, DNA assembly using *in vivo* homologous recombination in yeast represented an alternate method of combining DNA fragments into yeast centromeric shuttle vectors. The same PCR/restriction digest DNA products used for Gibson Assembly may be introduced into yeast according to the yeast transformation protocol described previously, with the resulting yeast colonies subjected to a screening colony PCR using primers designed to anneal with the insert. Candidate plasmids were then extracted from yeast, transformed into chemically competent *E. coli* and subjected to restriction digest DNA fingerprinting. Candidate plasmids displaying the

correct restriction pattern were then subjected to Sanger sequencing verification across junction regions.

2.7.16. 'Build a Genome' Protocol for De novo DNA Synthesis

DNA fragments were constructed *de novo* from overlapping oligonucleotides according to the 'Build a Genome' course protocol (Cooper *et al.*, 2012). This method is based on Polymerase Cycling Assembly (Stemmer *et al.*, 1995).

Synthetic DNA sequences were designed using the online GeneDesign web-based tool (Richardson *et al.,* 2012) using the "Building Block Design (Length Overlap method)" (Website: http://54.235.254.95/gd/). The web-based algorithm automatically generates the required overlapping oligonucleotides required for DNA synthesis. DNA fragments were designed as 'building blocks' of between 500 and 850 bp, with 40 bp regions of homology (if required) for Gibson Assembly. Each building block contained flanking non-internally cutting restriction sites so that building blocks can be simply excised from their backbone after TOPO cloning. Due to the error-prone nature of DNA synthesis, all DNA fragments were verified using Sanger sequencing.

Templateless PCR (T-PCR) reactions were undertaken by first diluting oligonucleotides to 300nM and combining 10 µL of each in a 200 µL total reaction volume ('templateless primer mix'). T-PCR reactions were prepared by combining 2 µL of 2.5 mM dNTPs with 5 µL of 5X Phusion GC buffer (NEB: New England Biolabs), 15.25 µL of ddH_2O, 0.25 µL of Phusion polymerase (NEB) and 2.5 µL of the templateless primer mix. Templateless PCR was then undertaken on a ProFlex PCR Thermal Cycler (Thermo Fisher Scientific), with settings according to **Table 2.8**.

Table 2.8: Thermal cycler program settings for templateless PCR.

	Cycle 1			Cycle 2		
	Denaturation	**Annealing**	**Elongation**	**Elongation**	**Annealing**	**Elongation**
Cycles	1			5		
Temperature	94°C	55°C	72°C	94°C	69°C	72°C
Time	3 min	30 s	1 min	30 s	30 s	1 min
	Cycle 3			**Cycle 4**		
	Denaturation	**Annealing**	**Elongation**	**Elongation**	**Annealing**	**Elongation**
Cycles	5			20		
Temperature	94°C	65°C	72°C	94°C	61°C	72°C
Time	30 s	30 s	1 min	30 s	30 s	1 min
	Finish					
	Elongation	**Hold**				
Cycles	1					
Temperature	72°C	8°C				
Time	3 min	Hold				

The Finishing PCR reaction was performed as follows. The T-PCR product was first diluted by a factor of 1 in 5 with ddH_2O. The 'Outer Primer Mix' (OPM) was prepared by mixing equal volumes of the first and last primers (6 uM) for a final concentration of 3 uM. Finishing PCR (F-PCR) was prepared by mixing 2 µL of 2.5 mM dNTPs with 5 µL of 5X Phusion GC buffer (NEB), 13.25 µL of ddH_2O, 0.3µL of Phusion polymerase (NEB) with 2 µL of the OPM and 2.5 µL of the diluted OPM. **Table 2.9** describes the finishing PCR reaction undertaken on a ProFlex PCR Thermal Cycler (Thermo Fisher Scientific). Once analysed on an agarose gel, DNA was subsequently gel excised, purified and TOPO cloned before subjected to Sanger sequencing verification.

Table 2.9: Thermal cycler program settings for finishing PCR.

	Cycle 1			Cycle 2		
	Denaturation	**Annealing**	**Elongation**	**Elongation**	**Annealing**	**Elongation**
Cycles	1			25		
Temperature	94°C	55°C	72°C	94°C	55°C	72°C
Time	3 min	30 s	1 min	30 s	30 s	1 min

	Finish	
	Elongation	**Hold**
Cycles	1	
Temperature	72°C	8°C
Time	3 min	Hold

2.7.17. Sanger DNA Sequencing

Sanger sequencing was performed on purified plasmid DNA using BigDye Terminator v3.1 Cycle Sequencing technology (Applied Biosystems) in 10 µL reactions. 2 µL of 5X sequencing buffer, 0.64 µL of primer (10 uM), 2 µL of BigDye and appropriate volumes of DNA (to a total of between 200 and 500 ng of DNA) was adjusted to 10 µL with ddH_2O. **Table 2.10** describes the thermal cycler program settings used for Sanger sequencing.

Table 2.10: Thermal cycler program settings for Sanger DNA sequencing.

	Amplification			Final Elongation	Cooling
	Denaturation	**Annealing**	**Elongation**		
Cycles	25			1	1
Temperature	95°C	50°C	60°C	60°C	4°C
Time	30 s	20 s	4 min	1 min 15 s	Hold

The reaction products were then delivered to Edinburgh Genomics (Edinburgh UK) for sequencing with sequencing trace files subsequently analysed using Snapgene software (Chicago, USA). Sequencing quality was determined using the standard Phred quality score i.e. a quality score of 30 represents a 99.9% base call accuracy.

2.7.18. Deep Sequencing

For deep sequencing and subsequent analysis, yeast samples were sent to Beijing Genomics Institute (Shenzhen, China). Sequencing was performed on a PGM Ion Torrent (Thermo Fisher Scientific).

2.7.19. Bioinformatics Methods

For DNA sequence manipulation (including design of vectors used to construct the neochromosome), restriction digest prediction, primer design, sequencing trace file analysis and vector maps used for figures in this study, Snapgene (Chicago, USA) software was used. Programming scripts were written using the Perl programming language. Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) was used to identify homology of neochromosome components to genome sequence databases located on NCBI (http://www.ncbi.nlm.nih.gov).

2.8. Specific Methods

The following materials and methods section describes specific methods related to Chapter 5 of this thesis.

2.8.1. RNA Extraction

RNA purification from yeast was performed according to the 'hot acid phenol' method and is based on the protocol described by Wei (2012) with differences noted as follows. Standard precautions were used for the preparation of RNA, including spraying surfaces and equipment with RNase *AWAY* (Molecular Bio Products) and the use of ddH_20 autoclaved with 0.1% diethylpyrocarbonate (DEPC) to remove RNases.

Overnight yeast cultures were re-inoculated into fresh media to a target OD_{600nm} of 0.3 and incubated with rotation to a final target OD of between 0.8 and 0.9 (i.e. early logarithmic phase). Cultures were centrifuged for 5 minutes at 2,103 g, washed with 10 mL of sterile ddH_2O, re-suspended in 1 mL of sterile ddH_2O and transferred to a 2 mL screw-cap centrifuge tube. Samples were centrifuged again (17,000 g for 1 minute), re-suspended in 500 µL of sterile NaAES buffer (50 mM sodium acetate (pH 5.3), 10 mM EDTA and 10% w/v SDS) before the addition of 500 µL of acid phenol solution (pH 4.3) and briefly vortexed. Samples were then incubated at 65°C for 30 minutes and chilled to room temperature on ice before centrifugation at 17,000 g for 20 minutes. The aqueous phase was transferred to fresh tubes containing equal volumes of phenol:chloroform:isoamyl alcohol (25:24:1), vortexing for 30 seconds, and centrifuging again at 17,000 g for 15 minutes. The aqueous phase was then transferred to fresh tubes containing equal volumes of chloroform, briefly vortexing, and centrifuging at 17,000 g for 10 minutes at 4°C. The final aqueous phase was transferred to tubes containing ice-cold (-20°C) absolute ethanol and precipitating at -20°C for 30 minutes plus a further -80°C for 10 minutes. Samples were then centrifuged at 17,000 g for 20 minutes at 4°C with the resulting RNA pellet washed with 500 µL of cold 70% v/v ethanol and centrifuging again for 10 minutes at 4°C. The resulting pellet was dried and re-suspended in 50 µL of RNase-free water, with RNA concentration quantified using the NanoDrop spectrophotometer (Thermo Scientific).

2.8.2. Polyacrylamide Gel Electrophoresis

Polyacrylamide gel electrophoresis (PAGE) was performed according to the method described by Rio *et al.* (2010) with differences noted as follows.

The PAGE gel was prepared by adding 24 g of urea, 1.66 mL of 3 M sodium acetate (pH 5.0), 13.33 mL of 30% acrylamide:bis (19:1), 5 mL of 10X TBE buffer (2.2 M Tris.HCl, 1.8 M boric acid and 50 mM EDTA), 0.5 mL of 10% (w/v) ammonia persulfate up to 50 mL with DEPC-treated water. Polymerisation was then induced by the addition of 50 µL of tetramethylethylenediamine (TEMED) solution and poured between clean glass plates. Once solidified, the PAGE gel was first pre-run in a vertical gel tank at 45 mA for 45 minutes with 1X TBE buffer. RNA samples were normalised to a total of 7.5 µg per well with equal volumes of 2X RNA loading dye (New England Biolabs) added to each sample. Prior to loading, all samples were heat-denatured by incubating at 95^{o}C for 5 minutes. Once the samples were loaded, the PAGE gel was run at 45 mA for around 3.5 hours.

2.8.3. Northern Blot: Electrotransfer

Electrotransfer of RNA onto a Hybond-N+ membrane (GE Healthcare Life Sciences) was performed according to an internal laboratory protocol. A Hybond-N+ membrane (GE Healthcare Life Sciences) was cut to size and placed facing the polyacrylamide gel while sandwiched between two sheets of Whatman 3MM blotting paper (GE Healthcare). Electrotransfer was performed in a vertical electrotransfer tank (EMBO) at 200 mA for 2 hours (constant voltage) at 4^{o}C before crosslinking with UV radiation. Nylon membranes were then wrapped in cling film and stored at 4^{o}C.

2.8.4. Northern Blot

Radiolabelling and Northern blot was performed according to internal laboratory protocols. Standard precautions were used when using radioactive γ^{32}P radionuclides.

The following probes were used:

- ***SUP61*****:** 5'-GCCCAAGAGATTTCGAGTCTCT-3'
  (Huang *et al.,* 2008)
- ***TRT2*****:** 5'-AATTGAACCCACGATCCCCGCATTACGAGTGCGATGCCTTACCACT-3'
  (Chakshusmathi *et al.,* 2003)
- ***TRR4*****:** 5'-ACTCGAACCCGGATCACAGCCACCGGAAGAATGCATGCTAACCATT-3'
  (Chakshusmathi *et al.*, 2003)
- ***5.8S*** **housekeeping control:** 5'-GCGTTGTTCATCGATGC-3'
  (Karkusiewicz *et al.*, 2011)

Radiolabelling of oligonucleotide probes was performed by combining 1.5 µL of 10X PNK buffer (New England Biolabs), 1 µL of the oligonucleotide probe, 2 µL of [γ^{32}P]-ATP, 1 µL of polynucleotide kinase (10 U/µL; New England Biolabs) and 9.5 µL of ddH_2O. Reactions were incubated at 37^{o}C for 1 hour on a heatblock, with probes purified using mini Quick Spin Columns (Roche Diagnostics) according to manufacturer's instructions.

Pre-hybridisation was performed by incubating the nylon membrane in hybridisation tubes containing 30 mL of hybridisation buffer (6X SSPE (900 mM NaCl, 60 mM sodium phosphate (pH 7.4), 6 mM EDTA), 5% Denhardt's solution (1% Ficoll (type 400), 1% polyvinylpyrrolidone 1% bovine serum albumin), 0.5% SDS in sterile ddH_20) at 37^{o}C for one hour. Hybridisation was performed by adding appropriate volumes of radiolabelled probe to hybridisation tubes and incubating overnight at 37^{o}C.

Membrane washes were performed with 30 mL of pre-warmed SSC buffer (20X is 3 M NaCl and 0.3 M sodium citrate), washing twice for 10 minutes at 37^{o}C. Salt concentrations of buffer were alternated depending on probe strength, with lower salt concentrations producing a stronger wash. Membranes were then exposed onto a photostimulated phosphor screen, with exposure time adjusted according to levels of radioactivity

measured on the membrane. Phosphorimager screens were then imaged on a Fujifilm FLA-5100 laser scanning system. Radioactive probes were stripped from nylon membranes by incubating at 37°C for 30 minutes repeatedly with 50 mL of near-boiling stripping buffer (0.1% SDS and 0.1X SSC). Probe stripping was verified by measuring for loss of radioactivity.

2.8.5. Neochromosome Linearisation

Telomere seed sequence integration and neochromosome linearisation was performed according to the method described by Mitchell and Boeke (2014). A specific description of the telomerator system can be found in Chapter 5.

Following the integration of the telomerator cassette into defined regions of the neochromosome, strains were transformed with a plasmid expressing the I-SceI homing endonuclease under a control of a *GAL1* promoter. Isolates were then inoculated into 10 mL of SC-His-Ura-Leu (2% glucose) and incubated with rotation at 30°C overnight. Once saturated, cultures were washed once with sterile PBS (phosphate buffered saline) solution and re-inoculated into 10 mL of SC-His-Leu containing either 2% galactose (for induction of I-SceI) and 2% glucose (negative control) to a target OD_{600nm} of 0.05 and incubated with rotation at 30°C for 24 hours. After 24 hours of growth, cultures were serially-diluted by a factor of 1 in 1,000 and plated onto SC-His + 5-FOA (1 mg/mL). Yeast strains cultured in glucose were observed to grow faster than those cultured in galactose – the approximate number of cells plated onto SC-His + 5-FOA was normalised to account for differences in cell concentration. Efficiency of linearisation was inferred by comparing the number of colonies of galactose-induced cultures with those cultured in glucose. After three days of growth, yeast isolates were verified using a colony PCR with primers designed to anneal with each arm of the telomerator cassette, but also to screen for the presence or absence of the junction region at the I-SceI cut site.

2.8.6. Pulsed-Field Gel Electrophoresis

Plug preparation and pulsed-field gel electrophoresis (PFGE) was performed according to a modified method described by Hage and Houseley (2013).

Preparation of agarose plugs was undertaken as follows:
Following two to three days of growth, 1 mL of cell culture was centrifuged, washed with 1 mL of wash buffer (10 mM Tris pH 7.6, 50 mM EDTA) and re-suspended in 50 µL of wash buffer containing 1 mg/mL lyticase. The cell suspension was pre-heated to 55^{o}C prior to the addition of 55 µL of molten agarose (SeaKem LE; pre-cooled to 55^{o}C). The agarose and yeast was then mixed thoroughly, added to PFGE molds (Bio-Rad) and allowed to set. Plugs were then added to 500 µL of wash buffer containing 1 mg/mL lyticase and incubated for three hours at 37^{o}C. Agarose plugs were then incubated overnight at 55^{o}C in 500 µL of PK buffer (100 mM EDTA, 0.2% sodium deoxycholate, 1% sodium lauryl sarcosine) containing 1 mg/mL proteinase K. The following day, plugs were washed with 1 mL of wash buffer four times, each by incubating at room temperature for 20 minutes with gentle agitation. Plugs were then stored at 4^{o}C in 500 µL of wash buffer.

Restriction digests of PFGE plugs were performed as follows:
Agarose plugs were first prepared as described above. For negative (circular) controls, plugs were cut in half and stored in 500 µL of wash buffer at 4^{o}C. The second half of plugs were then pre-conditioned by incubating in 200 µL of 1X Cutsmart restriction enzyme buffer (NEB) at 37^{o}C for 30 minutes. Agarose plugs were then subjected to restriction digest with I-SceI (NEB; 25 units per plug) in 200 µL of 1X restriction enzyme buffer (NEB) and incubated at 37^{o}C overnight, followed by a further I-SceI restriction enzyme digest (25 units per plug) for 3 hours at 37^{o}C. The plugs were then washed twice with 1 mL of wash buffer and stored in 500 µL of wash buffer at 4^{o}C prior to pulsed-field gel electrophoresis.

Pulsed-Field Gel Electrophoresis (PFGE) was undertaken as follows:
1.3% agarose gels were prepared by adding 1.56 g of Certified Megabase Agarose (Bio-Rad) to 120 mL of 0.5X TBE, heating to molten, and cooling to 60^{o}C. Agarose gels were then cast in a specialised casting tray and allowed to solidify. Agarose plugs were first cut to size and carefully loaded into the wells of the agarose gel. A thin strip of Lambda PFG Ladder (New England Biolabs) was cut with a scalpel and loaded where required. 1X TBE buffer was first circulated through the CHEF apparatus and pre-cooled to 12^{o}C.

Pulse-field running conditions can be altered depending on the size range of DNA. As the neochromosome is approximately 190 kb when linearised, the following settings were used (based on a recommended DNA separation of between 40-400 kb):

- **Gel percentage:** 1.3%
- **TBE:** 0.5X
- **Run time**: 24 h
- **Switch times (ramped):** 15–25 s
- **Voltage:** 6 V/cm

Once the run had finished, the agarose gel was first stained with 200 mL of 1X TBE buffer containing 1X SYBR Safe for 1 hour with gentle agitation, before being washed once with 200 mL of 1X TBE buffer for 20 minutes. The resulting gel was imaged using a Fujifilm FLA-5100 laser scanning system.

2.8.7. Neochromosome Extraction and Transfer

Neochromosome extraction and transfer was undertaken according to a modified protocol by Noskov *et al.* (2011). A further description of this method can be found in Chapter 5.

Overnight yeast cultures were re-inoculated into 400 mL of selective media and incubated at 30°C with shaking till optical density reached around 1.5. The resulting culture was split into two sets of 200 mL, centrifuged (1,860 g) and washed with 50 mL of sterile ddH_2O before re-suspension in 40 mL of SPE solution (1 M sorbitol, 0.01 M sodium phosphate, 0.01 M Na_2EDTA (pH 7.5)). 400 µL of Zymolyase 20T (10 mg/mL) and 40 µL of 2-mercaptoethanol was then added to each tube and incubated at 37°C for one hour. Spheroplasts were then pelleted by centrifuging at 1,294 g for 10 minutes at 4°C and gently re-suspended in 1 mL of 1 M sorbitol. 20 mL of high-pH lysis buffer was then added to the cells (0.05 M Tris-HCl, 0.02 M EDTA, 1% SDS (pH 12.8)), gently mixed, before the addition of 20 mL of phenol:chloroform:isoamyl alcohol (25:24:1) and mixing again. Samples were centrifuged at 4,194g for 20 minutes at room temperature with the supernatant transferred to fresh tubes. DNA was precipitated with 20 mL of isopropanol containing 2 mL of 3 M sodium acetate and centrifuged at 4,194 g for 30 minutes at 4°C with the resulting supernatant removed. The pellet was washed with 10 mL of 70% v/v ethanol, centrifuged again (4,194 g) to remove residual ethanol, and stored at -20°C prior to further purification.

The Noskov *et al.* (2011) protocol describes the purification of DNA using the Large Construct Kit (Qiagen). Due to significant similarities with the buffers and solutions used, the Plasmid Maxi Kit (Qiagen) was used instead, with minor alterations noted as follows (i.e. the QS buffer is found exclusively in the Large Construct Kit and was freshly prepared). Samples were re-suspended in 1 mL of TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8) before the addition of 3 µL of RNase A solution (10 µg/mL) and incubation at 37°C for 30 minutes. At this stage, duplicate samples were pooled together before exonuclease digestion using PlasmidSafe (Epicentre). 1.6 mL of ddH_2O, 80 µL of ATP solution (25 mM), 200 µL of 10X PlasmidSafe buffer and 40 µL of PlasmidSafe enzyme was added to the pooled sample and incubated at 37°C for 2.5 hours. 5 mL of QBT solution (Qiagen) and 10

mL of QS solution (1.5 M NaCl, 100 mM MOPS, 15% v/v isopropanol) was added to each sample before purification using MaxiPrep columns (Qiagen) using instructions described in the Qiagen Large Construct Kit. DNA samples were then stored at 4°C for a maximum of four days prior to yeast spheroplast transformation.

2.8.8. Neochromosome Plate Reader Growth Assay

To characterise yeast growth rate of strains housing variants of the tRNA neochromosome, the Tecan Sunrise microplate incubator/spectrophotometer was used. Yeast isolates were inoculated into selective media and incubated with rotation at 30°C overnight. The following morning, cultures were re-inoculated into fresh media to a target OD_{600nm} of 0.1 and loaded in 200 µL volumes into a CELLSTAR (Greiner Bio-One) 96-well optical growth plate. The following plate reader settings were used:

- **Temperature:** 30°C
- **Optical reading:** OD_{595nm}
- **Kinetic cycles:** 240 (15-minute intervals)
- **Shaking duration:** 895 seconds

Approximate yeast doubling time was calculated using an online tool (http://www.doubling-time.com/compute.php) and was inferred by calculating the differences in logarithmic phase between initial optical density, final optical density and the overall time in hours between these two values.

Chapter 3: Designing the tRNA Neochromosome

3.1. Engineering Approaches Rationalise the Design Process in Synthetic Biology

Biological systems are inherently complex. A challenge long faced by synthetic biologists is how to resolve this complexity and increase predictability when designing and performing experiments.

The increased scale, ambition and risk when working with designer biological systems generally renders more 'traditional' approaches to experimental design inadequate. Synthetic biologists have long recognised the advantage in applying engineering principles (Giese *et al.*, 2013), including the Design-Build-Test(-Learn) cycle, to rationalise experimental approach. Applied synthetic biology concepts include, but are not limited to, orthogonality, modularity, standardisation, a hierarchy of abstraction and a distinct separation between the design and construction processes (Heinemann and Panke, 2006).

Design plays an important role in any engineering discipline, and is important in synthetic biology to increase predictability and control. This is also true of the neochromosome: as every single nucleotide of the neochromosome may be defined, a design-based approached is essential to rationalise its complex structure and reduce attendant risk.

This chapter focuses on the design of the tRNA neochromosome. As with any engineering project, a series of design principles were required to define the project's parameters. These design principles are described as follows and were intended to rationalise and simplify each stage of the design process, maximise neochromosome stability and ensure that the designer chromosome functions *in vivo*. These design principles also largely form the overall structure of this thesis chapter.

1. **Robustness (Orthogonality):** To maximise neochromosome stability, functional and non-functional genetic elements should be recovered from non-*S. cerevisiae* yeast genome sequences whenever possible.
2. **Structural Hierarchy:** To rationalise design and form the basis of a modular structure, the neochromosome sequence should consist of increasingly large annotated segments, from the small (tRNA cassettes), medium (tRNA arrays) to the large (a higher-order structure).
3. **Control:** Origins of replication and replication termination sites should be strategically placed to define the DNA replication profile and minimise the risk of polar replication stall events associated with tRNA genes.
4. **Automation (Computer-Aided-Design):** To eliminate human error, computer scripts should be used to assign tRNA genes to orthogonal flanking sequences from the yeasts *Ashbya gossypii* (*A. gossypii*) and *Eremothecium cymbalariae* (*E. cymbalariae*).
5. **Standardisation/Modularity:** The neochromosome should consist of modular components based on a standardised design, allowing future alterations to the design, if necessary. The systematic presence of *rox* recombination sites also represent a unique, in-built modular system.
6. **Re-Writing:** To study aspects of biology, tRNA genes were re-designed to lack introns.

Robustness, a defined structural hierarchy, control and standardisation/modularity are commonly used concepts in engineering. Additionally, computer-aided-design is a well-proven principle in traditional engineering, and so automated programming scripts were used to assign tRNA genes to orthogonal flanking sequences and concatenate these elements. Re-writing by removing tRNA introns is an ancillary specification designed to address a specific question of biological relevance.

The tRNA neochromosome plays an essential underpinning role for the Sc2.0 consortium. To ensure that the neochromosome performs its necessary function as part of this consortium, the design of the neochromosome incorporated a series of specifications that were defined by other members of the Sc2.0 consortium (notably Prof. Jef Boeke, NYU School of Medicine, USA) prior to the inception of this doctoral study, and intended to unify the design and aims of this project with those of the overall Sc2.0 consortium. Those design specifications were defined as follows:

1. The neochromosome should house *rox* recombination sites flanking each tRNA cassette for orthogonal SCRaMbLE using Dre recombinase.
2. The neochromosome should consist of orthogonal sequences where possible. These include the flanking sequences, origins of replication and replication termination sites.
3. 500 bp 5' and 20 bp (later increased to 40 bp) 3' tRNA flanking sequences should be obtained from *A. gossypii*.
4. The total number of tRNA genes in the tRNA neochromosome should match that of the wild-type *S. cerevisiae* genome (*i.e*. 275).
5. Orthogonal replication origins and termination sites should be positioned in a manner to ensure neochromosome replication and prevent polar replication fork collisions. The unidirectional nature of tRNA orientation relative to the direction of the replication fork only later became a requirement.
6. tRNA introns should be removed.
7. tRNA arrays housing tRNA genes from each respective *S. cerevisiae* chromosome should be available as a resource to Sc2.0 consortium members.

The above specifications describe general guidelines for tRNA cassettes and tRNA arrays, but not a specific design *per se. i.e.* a detailed neochromosome design did not exist prior to the inception of this doctoral study. My role in this study was to unify these design specifications and ultimately form a coherent and detailed design. Unless where

otherwise specified in this chapter and described in the above list, the work in this chapter describes my own contribution to design. These include decision and considerations implemented in the design process. For example:

- The splitting of tRNA arrays into sequences <10 kb,
- The recovery of flanking sequences from *E. cymbalariae*,
- The design of the higher-order structure, and,
- The discovery and requirement to remove unwanted features from tRNA flanking sequences.

My role was also to work with Mr. Isaac Luo (Edinburgh, UK) to ensure the programming scripts used to generate the neochromosome sequence also met design requirements. Finally, the conceptual interpretation of engineering principles represents my own rationalisation of the neochromosome design process.

3.2. Orthogonality of Neochromosome Sequence Elements

Orthogonality is a concept used in synthetic biology to independently separate the operation of a non-native system from the cell's own native machinery (An and Chin, 2009). The principle of orthogonality was one of the fundamental features of the neochromosome as it was important to reduce instances of homologous recombination with the host genome. If genetic sequences had been duplicated from the genome of *S. cerevisiae* to be used on the neochromosome, the risk of reincorporation or rearrangement due to homologous recombination would significantly increase. Therefore, the nature and content of the neochromosome sequence was taken into careful consideration to limit homology with the host genome where possible.

Genetic sequences of yeast strains closely related to *S. cerevisiae* were identified to both maintain function and reduce the risk of homologous recombination. The elements used

from non-*S. cerevisiae* yeast species are summarised in **Table 3.1**. These elements are also discussed in more detail throughout this chapter.

Table 3.1: Summary of orthogonal genetic elements used for the neochromosomes. The column on the left summarises each orthogonal genetic element required, and the column on the right describes the yeast species used to recover these elements.

Genetic Element	Yeast Species
5' and 3' tRNA gene flanking sequences	*Ashbya gossypii* and *Eremothecium cymbalariae*
Origins of replication (Autonomously Replicating Sequences; ARS elements)	*Candida glabrata* (*C. glabrata*)
Replication termination (*Ter*) sites	*Saccharomyces paradoxus*, *Saccharomyces uvarum*, *Saccharomyces mikatae*, *Saccharomyces pastorianus*, and *Saccharomyces kudriavzevii*

3.3. Neochromosome Structural Hierarchy

A hierarchy of abstraction is an engineering attribute commonly used to manage complexity and define the information boundaries of complex systems (Heinemann and Panke, 2006). A central feature of the neochromosome was a defined structural hierarchy, which was designed to consist of components of increasing size, from the small (tRNA cassettes), medium (tRNA arrays) to the large (higher-order structure). This structural hierarchy was intended to rationalise the complexity of the neochromosome from a design perspective, form a basis for its modular structure and assist with the design of the defined replication profile described in Section 3.4. The hierarchical structure of the tRNA neochromosome is illustrated in **Figure 3.1**.

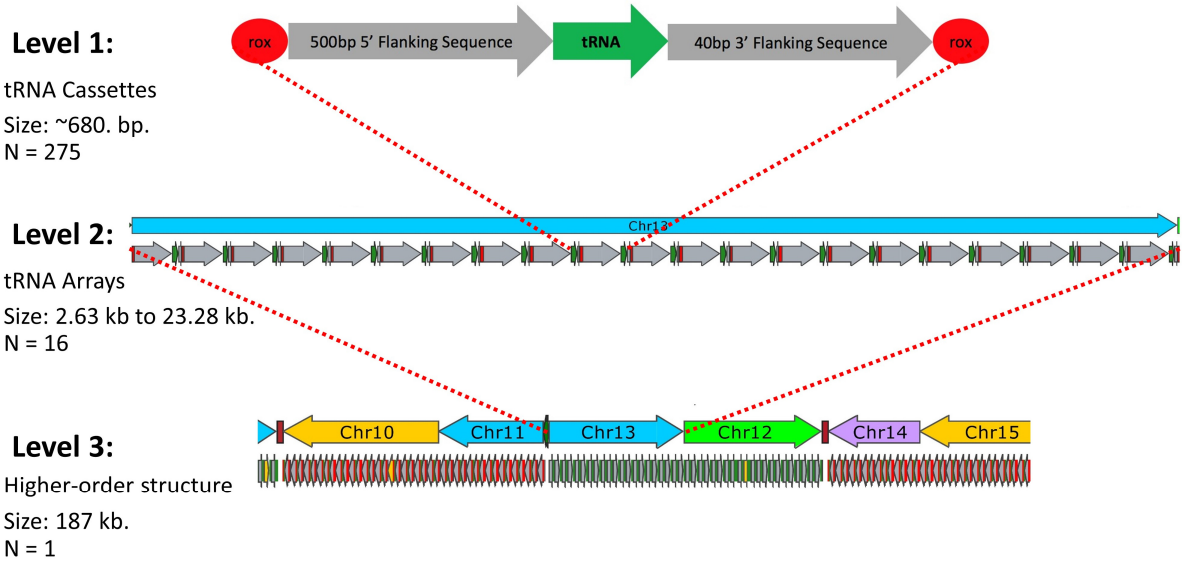


Figure 3.1. Hierarchical structure of the tRNA neochromosome. The figure is intended to illustrate the three levels of neochrochromosome structural hierarchy – the dotted red lines show that each component is part of a larger structure. In the above figure, 'N' indicates the number of each component (the neochromosome consists of 275 tRNA cassettes, 16 tRNA arrays and 1 higher-order structure). In this instance, the Chr13 tRNA array (blue arrow; level 2) consists of 21 tRNA cassettes (level 1), and lies adjacent to the Chr11 tRNA array and the Chr12 tRNA array on the higher-order structure of the tRNA neochromosome (level 3). For the final level of structural hierarchy (level 3), only part of the neochromosome is shown – the full neochromosome design may be observed in **Figure 3.4**.

3.3.1. Structural Hierarchy Level 1: tRNA Cassettes with Orthogonal Flanking Sequences

The smallest level of structural hierarchy represents the individual ~680 bp tRNA cassettes, each designed to individually house each of the 275 tRNA genes recovered from the *S. cerevisiae* genome. To maintain the functionality of each tRNA gene, each tRNA cassette was designed to incorporate important designer features (illustrated in **Figure 3.2**).

Surrounding each tRNA gene are 5' and 3' flanking sequences. A 500 bp 5' flanking sequence was incorporated to maintain a relative distance between each heavily-transcribed tRNA gene, and the 40 bp 3' region was incorporated to contain the necessary stretch of five or more poly-thymidine residues essential for transcriptional termination (Braglia *et al.*, 2005, Allison and Hall, 1985). As tRNA promoters are internal to the tRNA gene sequence itself (Schramm and Hernandez, 2002), flanking sequences from these yeast species were not expected to adversely affect tRNA expression.

tRNA flanking sequences cannot simply be recovered from the genome of *S. cerevisiae* as doing so would introduce significant homology between the neochromosome and the native genome, increasing the likelihood of genomic integration or *vice versa*. In addition, flanking sequences of tRNA genes in *S. cerevisiae* contain numerous transposable elements which preferentially associate with tRNA genes and generate highly-homologous solo LTR sequence repeats (Kim *et al.*, 1998). Therefore, to maintain the principle of orthogonality, provide structural stability and ensure tRNA functionality, the flanking sequences of tRNA genes were recovered from the yeast genome sequences of *A. gossypii* and *E. cymbalariae.* Because ensuring tRNA gene function takes priority over orthogonality, tRNA genes were directly obtained from the genome of *Saccharomyces cerevisiae*.

A. gossypii is a filamentous fungus originally isolated as a cotton pathogen, and now commonly used in industry as an overproducer of vitamin B2 which protects its spores against ultraviolet light (Stahmann *et al.*, 2000). The reason this species was originally selected for the recovery of tRNA flanking sequences is due to its exceptionally close genetic similarity and synteny with that of *S. cerevisiae,* and a lack of transposable elements (Dietrich *et al.*, 2013). Around 90% of all genes in *A. gossypii* are orthologous and syntenic with that of *S. cerevisiae*, and both species share a common genetic ancestor (Dietrich *et al.*, 2004).

However, *A. gossypii* houses only 199 tRNA genes (Dietrich *et al.*, 2004), and so lacks the required number of tRNA flanking sequences to match the 275 in *S. cerevisiae*. A second yeast species, *E. cymbalariae*, was identified due to its close genetic similarity to *A. gossypii*: 97% of all protein-coding genes in *E. cymbalariae* share a homologue with that of *A. gossypii* (Wendland and Walther, 2011). This species contains 143 tRNA genes and was identified to produce flanking sequences for the remaining tRNA genes in *S. cerevisiae.* Finally, *rox* recombination sites were designed to be incorporated both upstream and downstream of each 3' and 5' flanking sequence, respectively, for systematic chromosomal rearrangement. The automation of tRNA gene/flanking sequence assignment is described in Section 3.5, and the Dre-*rox* recombination system is described in Section 3.6.

Figure 3.2: tRNA cassettes with orthogonal flanking sequences and *rox* recombination sites. The figure is a representative example of the design used for each of the 275 tRNA cassettes of the tRNA neochromosome. The 500 bp 5' and 40 bp 3' flanking sequences are shown as grey arrows, the tRNA gene is indicated as a green arrow and *rox* recombination sites are shown as red circles.

3.3.2. Structural Hierarchy Level 2: tRNA arrays

The second level of neochromosome structural hierarchy are the individual tRNA arrays, each designed to house tRNA genes from each of the respective sixteen *S. cerevisiae* chromosomes. These tRNA arrays vary in size from 2.63 kb for the Chr1 tRNA array, to 23.28 kb for the Chr7 tRNA array (see **Table 4.2** in Chapter 4 for a full list by size).

Each tRNA array consists of tRNA cassettes that were concatenated using programming scripts (described in Section 3.5) so that relative tRNA gene orientation is unidirectional for each tRNA array, and that relative gene order is the same as that found in the *S. cerevisiae* genome. The unidirectional nature of tRNA orientation was designed in this manner to reduce the risk of polar replication fork stall events (described in Section 3.4.).

To increase the turn-around times for DNA synthesis, tRNA arrays were manually designed to be split into sections less than 10 kb. Each section was designed to contain a central 100 bp overlapping region, so that each tRNA array piece may later be seamlessly combined (see Chapter 4). These tRNA arrays, once introduced into pRS413 shuttle vectors, were intended to act as a resource for the Sc2.0 consortium to complement the loss of tRNA genes in each synthetic yeast chromosome. Finally, each tRNA array was designed to house flanking restriction sites to allow removal of each from their growth vector. **Figure 3.3** illustrates the Chr11 tRNA array in a pRS413 plasmid as a representative example.

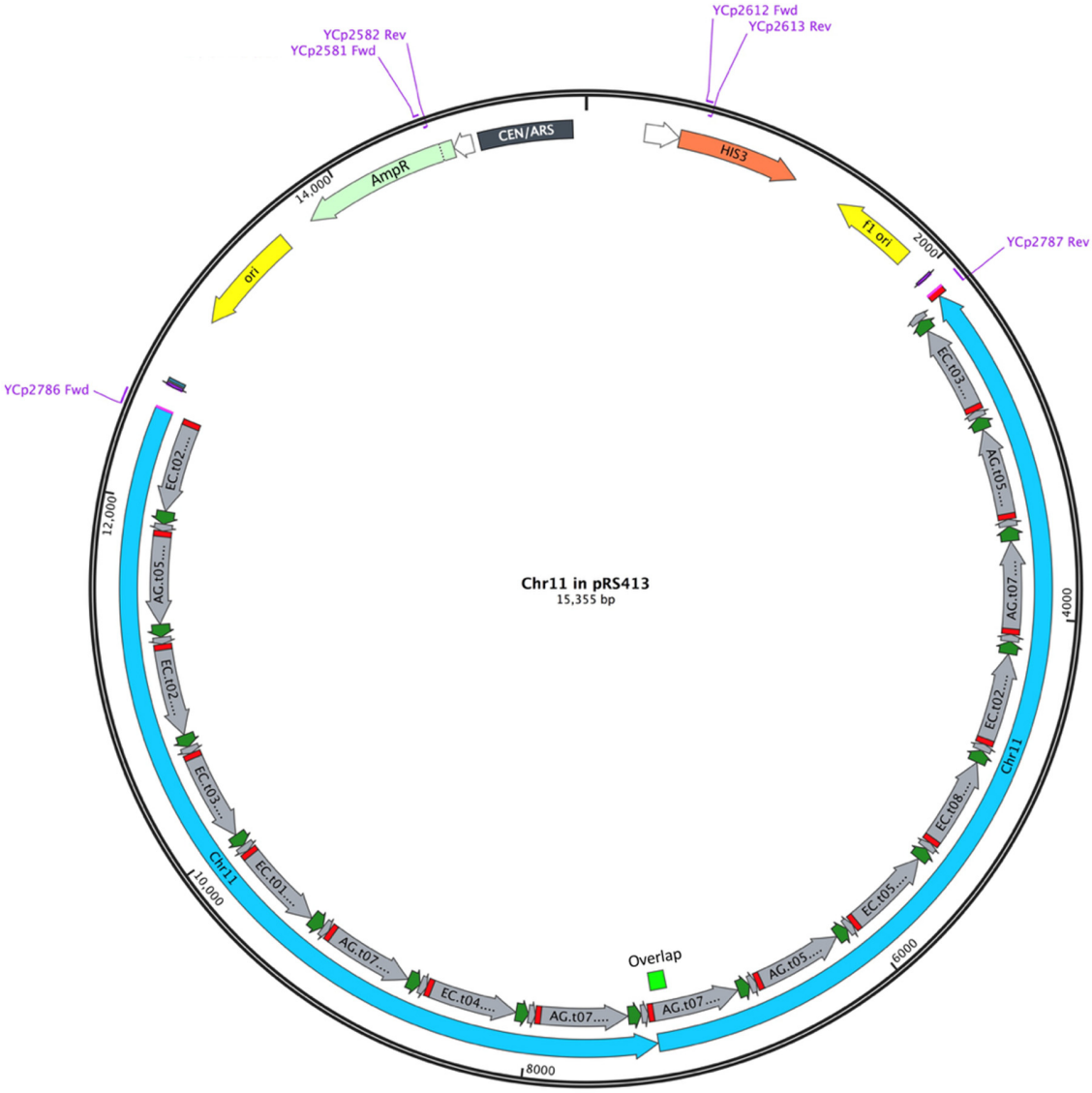


Figure 3.3: Representative plasmid map of the Chr11 tRNA array in pRS413. The figure provides an illustration depicting the common features observed for all other tRNA arrays. This tRNA array (15,355 bp) was synthesised in two sections (blue arrows; 5,201 bp and 5,136 bp, respectively), and later combined using Gibson Assembly (see Chapter 4 for further details). The central overlapping region is indicated in green, with primers used to amplify the pRS413 backbone indicated by the pink lines and numeric primer numbers. tRNA cassettes are indicated by grey arrows (tRNA flanking sequences), green arrows (tRNA genes) and red blocks (*rox* recombination sites). Flanking the tRNA array sequence lie NotI restriction sites (not shown), facilitating removal of the tRNA array from the pRS413 backbone.

3.3.3. Structural Hierarchy Level 3: Higher-Order structure

The final level of neochromosome design hierarchy is a defined, higher-order structure. This top-down approach to design was intended to incorporate the individual tRNA arrays into a true, functional chromosome, and include the necessary features for eukaryotic chromosome function *in vivo*. These features include the centromere, telomeres and origins of replication. The top-down design of the neochromosome (prior to linearisation) is indicated in **Figure 3.4** and was manually designed using Snapgene software (Chicago, USA). This was undertaken by manually positioning tRNA arrays in a manner intended to maintain relative inter-origin distance and ensure that tRNA genes face away from the leading edge of the replication fork to minimise replication stress (see Section 3.4 for further details)

Centromeres are regions of eukaryotic chromosomes where microtubules attach, *via* the kinetochore, to ensure equal segregation during mitosis. As it is unclear if an orthogonal centromere region would function in *S. cerevisiae*, the centromere and surrounding elements of pRS413 (the '*LEU2* inchworm vector') were identified to form the neochromosome backbone due to its characterised high mitotic stability (Hieter *et al.*, 1985, Murray and Szostak, 1983). The additional presence of an origin of replication next to the centromere region also ensures the ability to replicate during initial stages of construction.

A second defining feature of eukaryotic chromosomes are the telomeres. The presence of telomeres on smaller DNA sequences likely confers instability (Dani and Zakian, 1983), so it was decided to construct the neochromosome as a gradually increasing circular vector (a megaplasmid) prior to linearisation using the telomerator system (Mitchell and Boeke, 2014). The telomerator system is described in further detail in Chapter 5.

Introducing a sole DNA sequence of ~190 kb DNA in length into a yeast cell would not result in successful replication without the presence of replication origins. Although the centromere sequence on the neochromosome contains a single origin of replication, this was considered insufficient to replicate the full neochromosome sequence without a significant delay. Orthogonal ARS elements (kindly supplied by Prof. Conrad Nieduszynski (University of Oxford, UK) were recovered from the genome sequence file of *C. glabrata,* and contain Origin Recognition Complex (ORC)-binding motifs that closely resembles those found in *S. cerevisiae*. These origins include *chrF-444*, *chrL-615* and *chrM-794,* and were incorporated to be spaced at intervals of between ~38 kb to ~50 kb. Each was first shown to support plasmid replication in *S. cerevisiae* (Chapter 5).

A series of four synthetic, orthogonal replication termination sites were designed by Prof. Jef Boeke (NYU Langone Medical Centre, New York, USA) containing *Ter* sites from the genome sequences of *Saccharomyces paradoxus*, *Saccharomyces uvarum*, *Saccharomyces mikatae*, *Saccharomyces pastorianus* and *Saccharomyces kudriavzevii* genomes. These elements were designed to contain conserved Fob1 recognition sites, and stall DNA replication in both directions (i.e. act as dual replication fork blocks). These *Ter* sites were placed at spaced intervals in the neochromosome to terminate replication at defined regions (discussed in Section 3.4 and assayed in Chapter 5). Each *Ter* sites was designed as follows:

1. ***Ter1*:** A hybrid between the *TER2* and *TER1* sites of *Saccharomyces paradoxus* and the *TER1* and *TER2* sites of *Saccharomyces uvarum*.
2. ***Ter2:*** A hybrid between the *TER2* and *TER1* sites of *Saccharomyces kudriavzevii* and the *TER1* and *TER2* sites of *Saccharomyces pastorianus*.

3. ***Ter3:*** A hybrid between the *TER2* and *TER1* of *Saccharomyces mikatae* and the *TER1* of *Saccharomyces paradoxus* with a modified *TER2* hybrid containing a flipped middle spacer region.
4. ***Ter4:*** A hybrid between the *TER2* of *Saccharomyces paradoxus* and the *TER1* site of *Saccharomyces uvarum* as well as the *TER1* of *Saccharomyces kudriavzevii* and *TER2* of *Saccharomyces pastorianus* (modified with some minor alterations).

Figure 3.4: Overall circular structure of the tRNA neochromosome prior to linearisation. The overall size of the tRNA neochromosome is 186,802 bp. The centromere region of pRS413 and associated elements is indicated at the top of the sequence map, tRNA arrays (n = 16) are indicated by arrows of different colour, tRNA cassettes by small arrows (n = 275), *Ter* cassettes are indicated by red blocks (n = 4) and origins of replication (ARS elements) are indicated by green arrows (n = 3, with the fourth associated with the centromere). The above diagram was used as the final design for neochromosome construction.

3.4. Defining the Neochromosome Replication Profile to Minimise Polar Replication Fork Stall Events

The bespoke nature of the neochromosome presents an opportunity to control aspects of DNA replication and define the orientation of tRNA genes relative to the replication fork. A defined replication profile formed a central design consideration to minimise replication stress and instability associated with tRNA genes.

As described in Chapter 1, replication fork stall events are polar relative to the orientation of tRNA transcription. To reduce the risk of replication stress, tRNA genes of the neochromosome were orientated to face away from the leading edge of the replication fork. **Figure 3.5** illustrates the replication profile of the tRNA neochromosome, with blue and yellow arrows denoting the orientation of tRNA genes relative to replication origins.

Replication termination (*Ter* sites) were also positioned at regions of approximately equal inter-origin distance to minimise replication stress. These dual-block *Ter* sites were designed to terminate the replication fork in both directions at defined regions, and emulate their protective role in the rDNA locus by ensuring the replication fork faces the same orientation as tRNA transcription. *Ter* sites were designed to ensure that the replication fork does not progress beyond defined regions and collide with tRNA genes of opposing orientation. Finally, the presence of *Ter* sites was also intended to prevent two opposing replication forks colliding with each other.

Both origins of replication and *Ter* sites were spaced in a manner to minimise inter-origin distance and ensure that no individual region encounters an extensive chronological delay during replication.

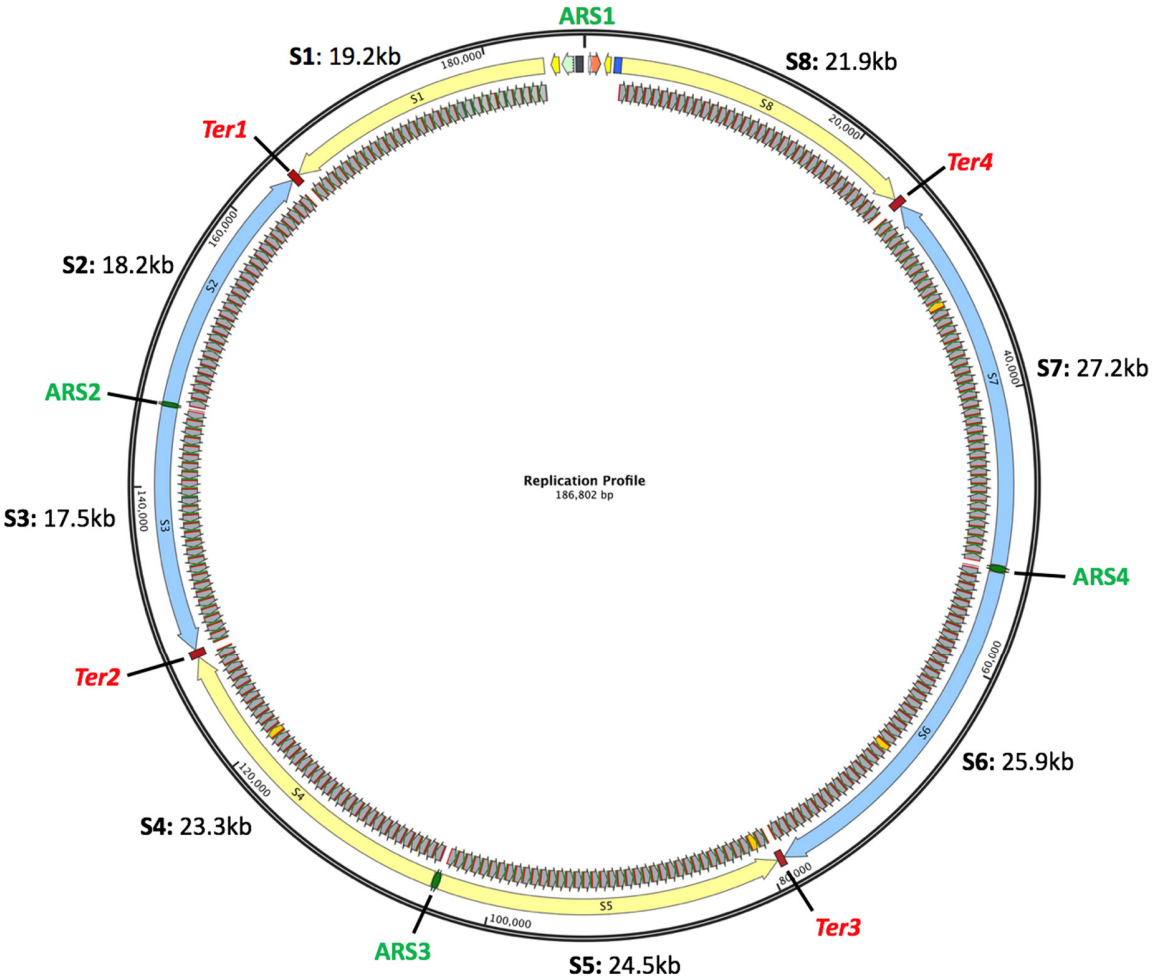


Figure 3.5: Graphical overview of the neochromosome replication profile. The figure illustrates the overall design intended to ensure even replication and minimise polar replication stall events. Yellow and blue arrows show the direction of each respective replication fork, with segment sizes (S1 to S8) indicated. Origins of replication are shown in green, and replication termination sites are shown in red. ARS1 is the origin associated with the centromere and replicates segment 1 and segment 8, ARS2 is *chrF-444* and replicates segment 2 and 3, ARS3 is *chrL-615* and replicates segment 4 and 5 and ARS4 is *chrM-794* and replicates segment 6 and 7. tRNA genes are specifically designed to orientate away from the leading edge of the replication fork to minimise polar replication fork collisions. The presence of *Ter* sites ensure that DNA replication terminates at defined regions.

3.5. Automation of Neochromosome Sequence Concatenation

Process automation and computer-aided design are commonly used tools in engineering. A series of programming scripts were used to automate the process of assigning tRNA genes to flanking sequences and generate each tRNA array sequence. This approach has several advantages compared to manually designing the neochromosome: first, automation greatly reduces the inherent risk of human error; second, fundamental design changes can be undertaken in seconds; and third, the speed of sequence concatenation is orders-of-magnitude faster than a manual approach.

An algorithm was designed to match tRNA genes from *S. cerevisiae* to flanking sequences recovered from the genome sequences of *A. gossypii* and *E. cymbalariae* (**Figure 3.7**). These programming scripts were designed to generate an available 'pool' of sequence elements that were concatenated to form tRNA cassettes and, subsequently, tRNA arrays. However, some tRNA flanking sequences contained undesirable features (summarised in **Figure 3.6**), and were modified using automated scripts or otherwise manually discarded from the available pool altogether (see Sections 3.5.1 to 3.5.4). 55% and 60% of all tRNA genes in *A. gossypii* and *E. cymbalariae*, respectively, had unwanted features in their 5' flanking sequence. These include neighbouring gene starts, neighbouring tRNA genes and, for *E. cymbalariae*, solo LTR elements.

Programming scripts were based on a series of design specifications to ensure that the neochromosome met project parameters. These design specifications are described as follows:

1. tRNA genes of *S. cerevisiae* should match flanking sequences in *A. gossypii* and *E. cymbalariae*, in that order of preference, based on relative orientation and matching anticodon where possible.

2. Flanking sequences should only be assigned to tRNA genes once.
3. tRNA genes will lack introns.
4. All 3' tRNA flanking sequences should possess five or more thymidine residues required for transcriptional termination.
5. The 500 bp 5' flanking sequences should not contain unwanted features. These include genes, solo LTR elements and other tRNA genes.
6. *rox* recombination sites will flank each tRNA cassette.

It should be noted that the Chr6 tRNA array was manually designed prior to the development of programming scripts, but was used as a starting point for the above design specifications. Except for shorter 20 bp 3' flanking sequences, no discrepancies in design specification were observed for the Chr6 tRNA array, and no differences were observed between the manually-designed Chr6 tRNA array and its computer-generated counterpart.

All programming scripts used to generate tRNA array sequences were written by Mr. Isaac Luo (Edinburgh, UK) based on the Python programing language with BioPython modules. Isaac Luo and Dr. Patrick Cai (Edinburgh, UK) were also largely responsible for the development of the algorithm used to assign *S. cerevisiae* tRNA genes to flanking sequences (**Figure 3.7**). Finally, features associated with the implementation of programming scripts (detailed in the following two paragraphs), are largely the work of Mr. Luo and Dr. Cai.

S. cerevisiae tRNA gene sequences (lacking introns), and *A. gossypii*/*E. cymbalariae* 3' and 5' flanking sequences were recovered from Genbank files downloaded from the NCBI genome database (accession numbers are found in Appendix III), with metadata exported to comma separated value (.CSV) files using programming scripts. All tRNA gene anticodons were determined (*i.e.* independently verified) by 'calling' a program compiled

from the source code of tRNAscan-SE (Schattner *et al.*, 2005). Following tRNA gene/flanking sequence assignment, tRNA cassettes were concatenated into tRNA arrays and generated as Genbank output files. To validate the reliability of computer scripts, sequences were manually spot-checked at random to ensure the neochromosome met the above design principles.

Finally, to simplify annotation, a naming convention was developed for tRNA genes in *S. cerevisiae*, *A. gossypii* and *E. cymbalariae.* Each sequence (and hence their corresponding flanking sequences) were assigned a unique ID based on relative tRNA order i.e. AG.t6.4 corresponds to the fourth tRNA gene on the 6th chromosome of *Ashbya gossypii*. Information retrieved using the above methods was also used to annotate sequences with corresponding metadata. For example, clicking on a sequence element in Snapgene (Chicago, USA) will reveal the design history of that element. This includes the species origin, tRNA anticodon and feature alterations performed on flanking sequences.

3.5.1. Silencing Unwanted Genes in 5' Flanking Sequences

Unwanted gene starts transcribed by RNA Pol II were occasionally observed to reside in the 5' flanking sequence upstream of tRNA genes in *A. gossypii* and *E. cymbalariae.* To prevent unwanted partial translation, genes found with an embedded start codon in the 500 bp 5' flanking sequences were silenced using programming scripts by altering the 'ATG' start codon to the 'AAG' lysine codon (**Figure 3.6A**). Embedded genes were unmodified (ignored) if their trailing 3' sequence was found within the 5' flanking sequence. 55 tRNA genes in *A. gossypii* and 26 in *E. cymbalariae* were observed to be affected in this manner.

3.5.2. Co-Transcribed tRNAs and Shared Flanking Sequences

Some tRNA genes in *A. gossypii* and *E. cymbalariae* were found to be co-transcribed i.e. tRNA genes appear together as groups of two and transcribed together from a common upstream tRNA promoter (Straby, 1988) (**Figure 3.6B**). To prevent the presence of non-*S. cerevisiae* tRNA gene in the 5' flanking sequence, the 5' flanking sequence of the downstream tRNA and the 3' sequence of the upstream tRNA were used. Twelve tRNA genes from *A. gossypii* and twelve from *E. cymbalariae* were found to be affected.

In rare incidences, *A. gossypii* and *E. cymbalariae* tRNA genes in close proximity, but facing opposite orientations, were found to share a portion of the same 5' flanking sequence (**Figure 3.6C**). To prevent homology when selecting each flanking sequences, one of the two tRNA genes was discarded based on relative need for the other tRNA anticodon. This scenario occurred twice in *A. gossypii* (AGOS_t0049 + t0050 and t0071 + t0074) and three times in *E. cymbalariae* (Ecym_7336 + 7337, 4463 + 4464 and 2337 + 2338).

3.5.3. Removal of Solo LTR Elements in E. cymbalariae *5' Flanking Sequences*

E. cymbalariae contains a single Ty-Gypsy-type transposable element adjacent to a tRNA gene (Ecym_5174). As transposable elements preferentially incorporate in the 5' upstream sequence of tRNA genes, several annotated solo LTR elements were observed to occupy these regions. *E. cymbalariae* was found to have approximately 50 solo LTRs in the 5' flanking sequences, or around 35% of the 144 total number. Scripts were designed to eliminate these solo LTR elements entirely, with the resulting 5' sequence extended to include the required 500 bp (**Figure 3.6D**). If a second solo LTR was observed in the extended sequence, it too was deleted. As the Ecym_5174 tRNA contains the transposable element, it was discarded from the available sequence pool.

3.5.4. Ensuring the Presence of Transcriptional Terminators

Poly-thymidine residues in the 3' flanking sequence are critical for successful termination of tRNA transcription. In *A. gossypii* and *E. cymbalariae*, 95% and 86.8% of all 3' tRNA flanking sequences had five-or-more thymidine residues located within 40 bp. The remaining tRNA genes were found to be either co-transcribed or, alternatively, misannotated in terms of orientation based on the presence of the poly-T residue appearing on the reverse complement strand (**Figure 3.6E**). These misannotated genes include one tRNA gene in *A. gossypii* (AGOS_t0191) and five in *E. cymbalariae* (Ecym_1186, 1246, 2286, 5513 and 2084). Additionally, *E. cymbalariae* had eight tRNA genes with the poly-T terminator appearing after 40 bp. Flanking sequences that lacked the terminator sequence were discarded from the available pool.

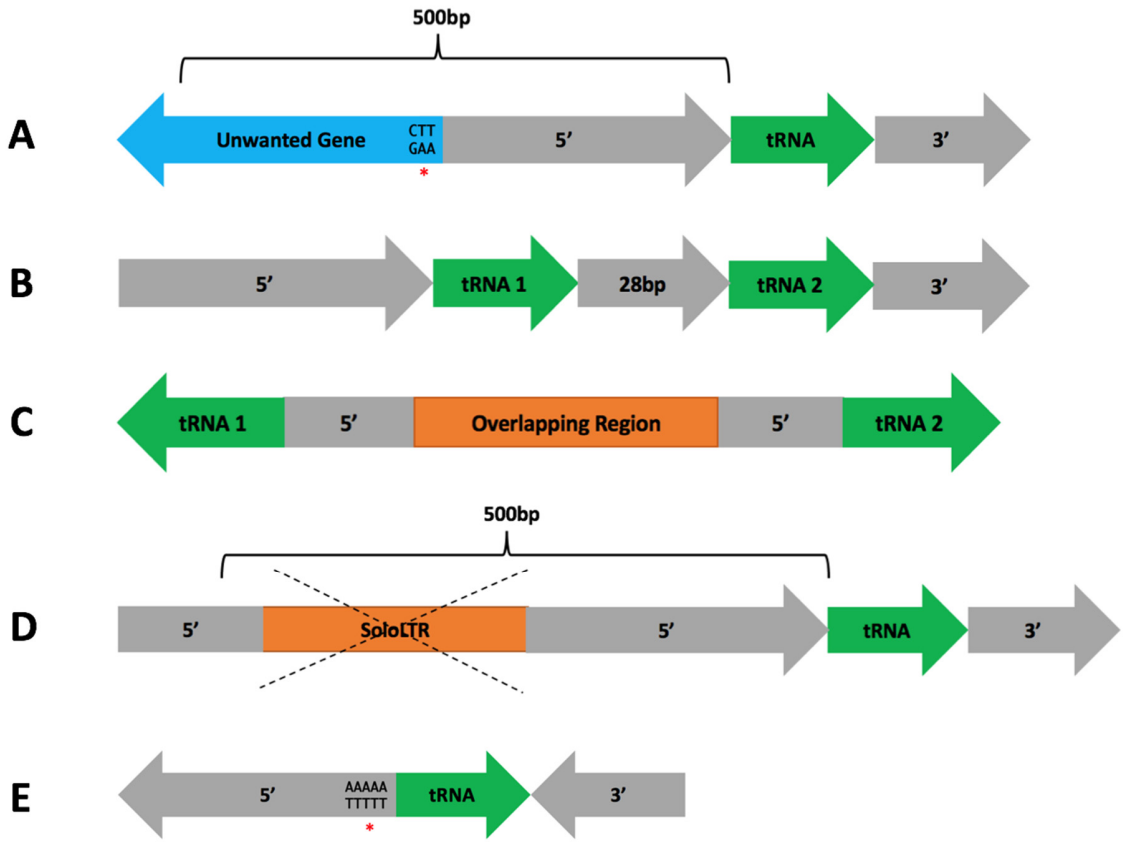


Figure 3.6: Graphical overview of the unwanted features identified in tRNA flanking sequences. The figure provides a graphical representation of each of the features altered in the tRNA flanking sequences of *A. gossypii* and *E. cymbalariae* prior to generating tRNA cassettes. (**A**) Unwanted genes in 5' flanking sequence. Partially transcribed gene 'starts' identified in the 5' flanking sequence were silenced by altering the 'ATG' start codon to the 'AAG' lysine codon (denoted by the red asterisk). (**B**) Co-transcribed tRNA gene pairs separated by a 28 bp spacer region. The 5' flanking sequence of tRNA 1 and the 3' flanking sequence of tRNA 2 were used. (**C**) Shared 5' flanking sequence between two divergent tRNA gene clusters. Only one of the two tRNA flanking sequences were used based on relative anticodon requirement. (**D**) Solo LTR repeat observed in 5' flanking sequences of *E. cymbalariae.* The solo LTR was removed, with the 5' flanking sequence extended to incorporate the full 500 bp. (**E**) Misannotated tRNA orientation. This was inferred by the presence of a poly-thymidine residue in the 5' flanking sequence (denoted by the red astetrisk) and its absence in the 3' flanking sequence. These tRNA flanking sequences were discarded from the available pool.

3.5.5. tRNA Gene Flanking Sequence Assignment: Algorithm Design

Following the removal of unwanted features in tRNA flanking sequences, an algorithm was designed to form the basis of assigning tRNA genes to their respective flanking sequences (**Figure 3.7**). The algorithm was designed to assign the tRNA genes of *S. cerevisiae* to flanking sequences in *A. gossypii* and *E. cymbalariae*, in that order of preference, based on relative orientation and matching anticodon. Of the 275 tRNA genes, the vast majority (240) were assigned flanking sequences in this manner. Of the remaining 35 tRNA genes, 19 were assigned flanking sequences based on matching amino acid while the remaining 16 tRNA genes were randomly assigned flanking sequences from tRNAs irrespective of anticodon or amino acid. **Table III** in Appendix IV details a full list of each *S. cerevisiae* tRNA gene and its associated assigned flanking sequence.

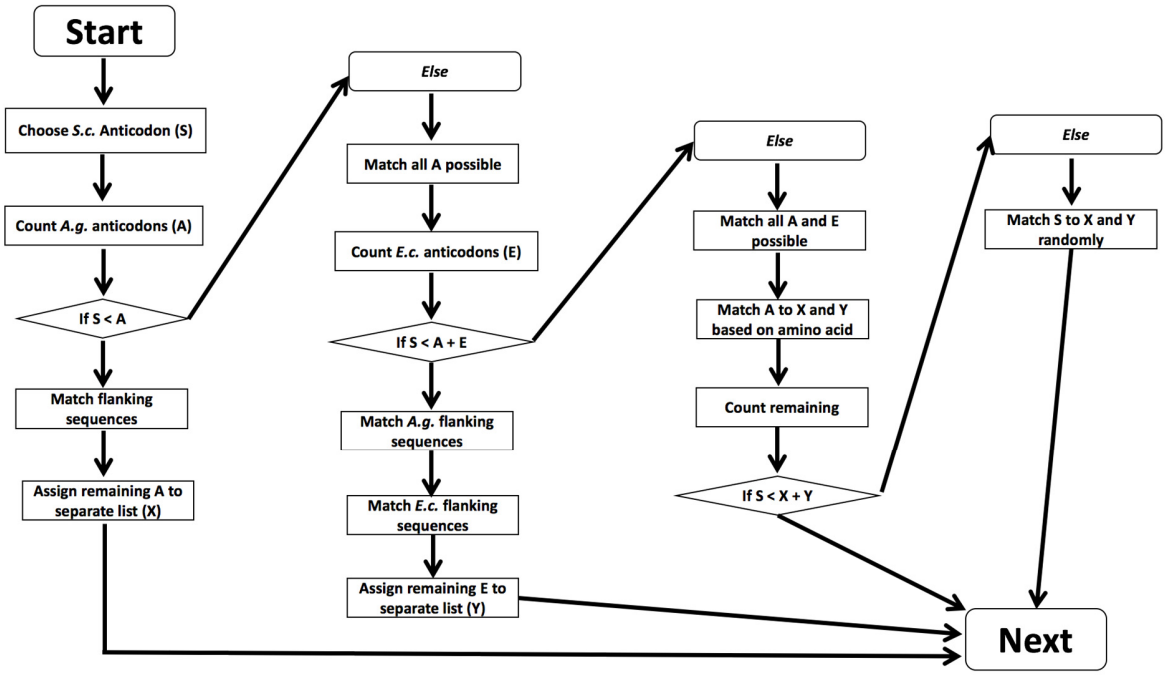


Figure 3.7: Algorithm used to assign *S. cerevisiae* tRNA genes to flanking sequences of *A. gossypii* and *E. cymbalariae.* The flow chart describes the overall process used to assign *S. cerevisiae* tRNA genes to flanking sequences of *A. gossypii* and *E. cymbalariae,* and ultimately form the basis of programming scripts used to generate the tRNA arrays of the neochromosome. Mr. Isaac Luo and Dr. Patrick Cai are largely responsible for the deisgn of the above algorithm. The first stage of the algorithm identified a single anticodon family of *S. cerevisiae* (**S**) and assigned as many *A. gossypii* (**A**) flanking sequences with the same anticodon as possible. The remaining unassigned flanking sequences were then added to a separate list (**X**). For tRNA genes that cannot be assigned *A. gossypii* flanking sequences based on this principle (IF **S**>**A**), the second stage of the algorithm performed the same process of assignment based on tRNA flanking sequences from *E. cymbalariae* (**E**) and unassigned flanking sequences were again added to a separate list (**Y**). For tRNA genes that cannot be assigned flanking sequences by matching anticodon from either of these two species (IF **S**>**A**+**E**), tRNA genes were assigned flanking sequences based on matching amino acid from the two *A. gossypii* and *E. cymbalariae* lists (**X** and **Y**). Finally, the remaining tRNA genes that cannot match any of these conditions were randomly assigned flanking sequences irrespective of anticodon, amino acid or species.

3.6. In-Built Modularity: Dre-*rox* SCRaMbLE

SCRaMbLE is a Cre-*lox*-based recombinase system designed to induce directed evolution and gain of function in synthetic yeast strains. The *de novo* design of the tRNA neochromosome presented an opportunity to introduce a second site-specific recombinase system. To prevent cross-interaction (and subsequent re-incorporation with the synthetic host genome), a heterospecific orthogonal system based on Dre recombinase and *rox* recombination sites was identified (Sauer and McDermott, 2004). There is no evidence of any cross-reactivity between the Cre-*lox* and Dre-*rox* systems (Anastassiadis *et al.*, 2009).

Flanking each tRNA cassette lie *rox* recombination sites (**Figure 3.2**). Inducing SCRaMbLE in the neochromosome context presents an opportunity to systematically study new aspects of tRNA genetics by altering tRNA copy number. For instance, inducing SCRaMbLE under different conditions, such as that of stress or nutrient starvation, may allow the determination if there is a preference for individual tRNA copy number optimal to yeast stressors. As the Cre-*lox* and Dre-*rox* systems do not cross-interact, both systems may also be activated at the same time ("Co-SCRaMbLE"), so that optimal tRNA anticodon arrangement may be optimised depending on the arrangement and content of the genome. Finally, the Dre-*rox* SCRaMbLE-based system may allow for the systematic incorporation of genes or metabolic pathways specifically into the neochromosome.

As the final synthetic yeast genome is yet to be constructed, the neochromosome was constructed in wild-type (and semi-synthetic) strain backgrounds. For this reason, much of the neochromosome is non-essential, and so inducing Dre-*rox* SCRaMbLE in the neochromosome context will likely lead to a loss of DNA without biological significance. The study of Dre-*rox* SCRaMbLE in the context of the tRNA neochromosome is not part of the scope of this doctoral study.

3.7. Re-Writing: Intron removal

One of the objectives of the Sc2.0 consortium is to study the biology of introns and so most introns are to be removed from the synthetic yeast genome (Dymond *et al.*, 2011). These include the introns of tRNA genes. As described in Chapter 1, the essentiality of introns in tRNA genes remains an open question. In *S. cerevisiae*, 21% of all tRNA genes have introns, while only 4.3% of all genes contain introns (283 introns in approx. 6607 genes). Therefore, tRNA genes (n = 275) account for 20.9% of the total number of introns, while only accounting for 4.2% of the total number of genes (approx. 6,607). Their presence appears to have some significance.

While tRNA introns are poorly conserved and poorly understood, an inherent risk is recognised in their systematic removal. As described previously, introns may have unknown functions including modification of tRNA products, although previous work has shown that it is possible to removal all introns from a single tRNA species without any apparent loss of function (Mori *et al.*, 2011). It is unknown if tRNA introns are essential, although their systematic removal will present an answer to this long-standing question and enable new opportunities to study their function in future.

3.8. Discussion

The tRNA neochromosome was successfully designed based on a series of principles informed by engineering concepts. These included orthogonality, a defined structural hierarchy, in-built modularity and re-writing (intron removal). The defined profile of neochromosome replication was an important feature to minimise the risk of replication stress associated with tRNA genes.

While neochromosome design was optimised based on the available components, design trade-offs were undertaken due to the current limitations of available components and DNA synthesis capabilities. For example, tRNA arrays, of varying size, could not be easily split into smaller sequences following DNA synthesis and so presented a constraint on the design of the higher-order structure. Secondly, it was unclear what the optimum arrangement and copy number of replication origins were required due to the hard-to-replicate nature of tRNA genes. Finally, a general lack of the number of dual-block *Ter* sites available restricted the available combinations of the neochromosome structure. If the number of replication origins were to increase, so too would the number of *Ter* sites required.

Although the process of matching tRNA genes to orthogonal flanking sequences was automated, significant user input was required to generate the final neochromosome sequence. Programming scripts reduce greatly the risk of human error when generating DNA sequences, although in turn introduce systemic risk caused by scripts not performing as expected. When developing scripts to produce tRNA arrays, 'bugs' manifested as incorrectly removed Solo LTR elements, genes in 5' flanking sequences incorrectly silenced, tRNA genes matched more than one pair of flanking sequences and even tRNA genes inadvertently obtained from *A. gossypii* instead of *S. cerevisiae*. A close working relationship between designer and programmer was therefore required to correct these scripts and ensure that each tRNA array sequence met the original design principles.

Overall, the above neochromosome design represents something of a starting point in the recursive Design-Build-Test-Learn cycle. Should more functional components become available to the designer, this cycle may be revisited in future with alterations and optimisations undertaken based on knowledge acquired during construction. The process of neochromosome construction is described in the following chapter (Chapter 4).

Chapter 4: Constructing the tRNA Neochromosome

4.1. Introduction

This chapter describes the construction of the tRNA neochromosome from its constituent tRNA arrays. At the time of writing, no technologies have been described that provide the means to construct synthetic eukaryotic neochromosomes *de novo,* and so this chapter describes the development of a method to do so. This work is also intended to provide synthetic biologists with a technology that can be applied to build neochromosomes in future.

The construction of a tRNA neochromosome introduces a series of challenges. The first includes translating the *in silico*-based sequence of the neochromosome design into a real-world stretch of DNA. The Build-a-Genome (BAG) protocol may have been used to assemble shorter DNA fragments *de novo*. However, due to the large size (186 kb) of the tRNA neochromosome, the BAG protocol would have been exceptionally time-consuming. Synthetic biologists now have the option to order DNA sequences from external commercial vendors, which greatly reduces the turnaround time between designing and synthesising DNA. However, limitations do exist: external vendors do not currently have the technology to synthesise DNA on the chromosome scale, and so DNA sequences must be split into smaller components for later assembly. To reduce the time required for DNA synthesis, the neochromosome was constructed from DNA fragments less than 10 kb in size.

A second challenge faced is *how* to construct the neochromosome from these smaller DNA fragments. Many current DNA assembly strategies, such as Gibson Assembly and Golden Gate, are limited by the size of DNA they can join together. However, large-scale DNA construction technologies do exist, and include the method described by Gibson *et*

al. (2010) who used *in vivo* homologous recombination in *S. cerevisiae* to construct the *Mycoplasma mycoides* genome. The great power of *in vivo* homologous recombination in yeast was identified as the primary means to construct the neochromosome using the *inchworming* method described in Section 4.2.2.

The initial section of this chapter describes the experimental approach used for each stage of neochromosome construction, from the combining of split tRNA arrays, assembly of vectors used to facilitate neochromosome construction to the process of construction and verification itself. Instead of detailing the construction of every single component associated with the neochromosome, the results section of this chapter includes a representative example of each stage of neochromosome assembly. A flow chart of the primary steps involved in neochromosome construction may be found in **Figure 4.1**.

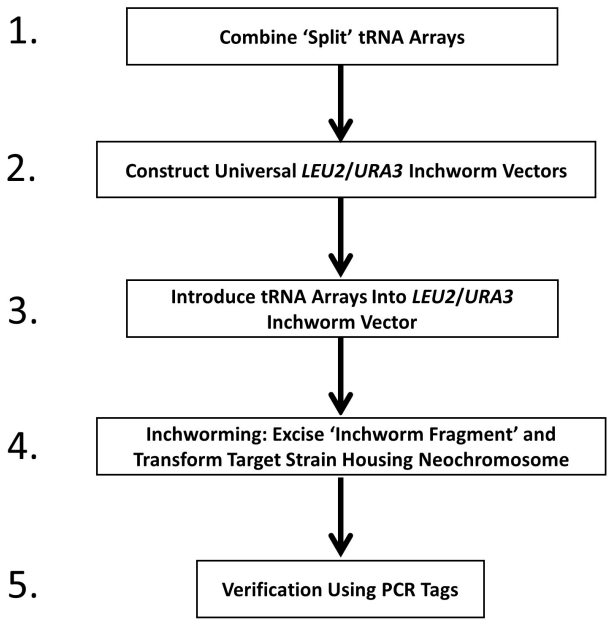


Figure 4.1: Flow chart illustrating the primary steps involved in neochromosome construction. Part 1 on the above flow chart is described in Section 4.2.1. Part 2 is described in Section 4.2.3, part 3 in Section 4.2.4, part 4 in Section 4.2.2 and 4.2.5 and part 5 is described in Section 4.2.5.

4.2. Experimental Approach

4.2.1. Combining 'Split' tRNA Arrays and De novo *Construction of the Chr1 tRNA Array*

To reduce the overall time taken to synthesise DNA fragments, tRNA arrays over 10 kb in size were designed using Snapgene software (Chicago, USA) to be split into sequences less than 10 kb. These included the Chr4, Chr5, Chr7, Chr10, Chr11, Chr12, Chr13, Chr15 and Chr16 tRNA arrays. A 100 bp overlapping region, unique to each tRNA array, was manually incorporated into the central region of each tRNA array to facilitate their subsequent assembly. **Figure 4.2** summarises the primary steps in the process of combining split tRNA arrays.

To combine each tRNA array part, the pRS413 backbone was first amplified using PCR with primer pairs designed to house 40 bp regions of overlapping homology. The *AmpR* and *HIS3* markers were designed to be split into two (indicated by 'A' and 'B' in **Figure 4.3**) to provide an additional means of selection for successful construction following Gibson Assembly or *in vivo* homologous recombination. These primers also contained flanking non-internally cutting restriction sites to subsequently recover each tRNA array from their backbone. Following recovery of each tRNA array part through restriction digest, the overlapping homology regions allowed each tRNA array part to be combined using either Gibson Assembly or *in vivo* homologous recombination in BY4741. The different strategies used to combine split tRNA arrays is summarised in **Table 4.1**.

In instances where Gibson Assembly failed (such as the three-piece 23 kb Chr7 tRNA array), assembly using homologous recombination in BY4741 was observed to be successful in every instance. After each tRNA array displayed the correct restriction digest fingerprint pattern following bacterial miniprep, junction regions of assembly were subjected to verification using Sanger sequencing. **Figure 4.3** provides a schematic

representation of this process and Section 4.3.1 describes a representative example of combining two halves of the Chr11 tRNA array.

The Chr1 tRNA array is the smallest of all sixteen tRNA arrays (2.6 kb) and was constructed by Ms. Lois Ogunlana (Edinburgh, UK) under my supervision using the Build a Genome protocol. This DNA sequence was split into four 'building blocks' containing 100 bp overlapping regions, combined using Gibson Assembly into a pRS413 backbone, and later verified using Sanger sequencing. The Chr1 tRNA array was observed to be sequence-perfect.

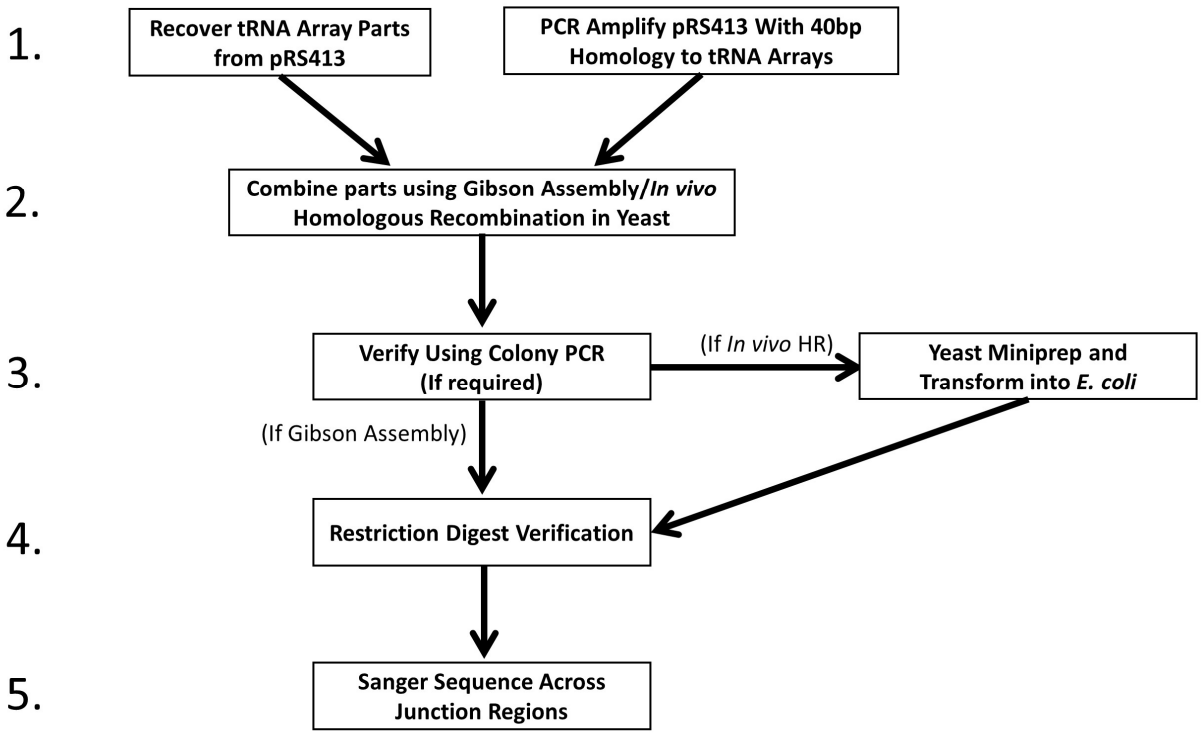


Figure 4.2: Flowchart summarising the primary steps involved in the process of combining split tRNA arrays. As described on step 3 on the above diagram, either Gibson Assembly or *in vivo* homologous recombination in yeast (*in vivo* HR) may be used to generate each construct.

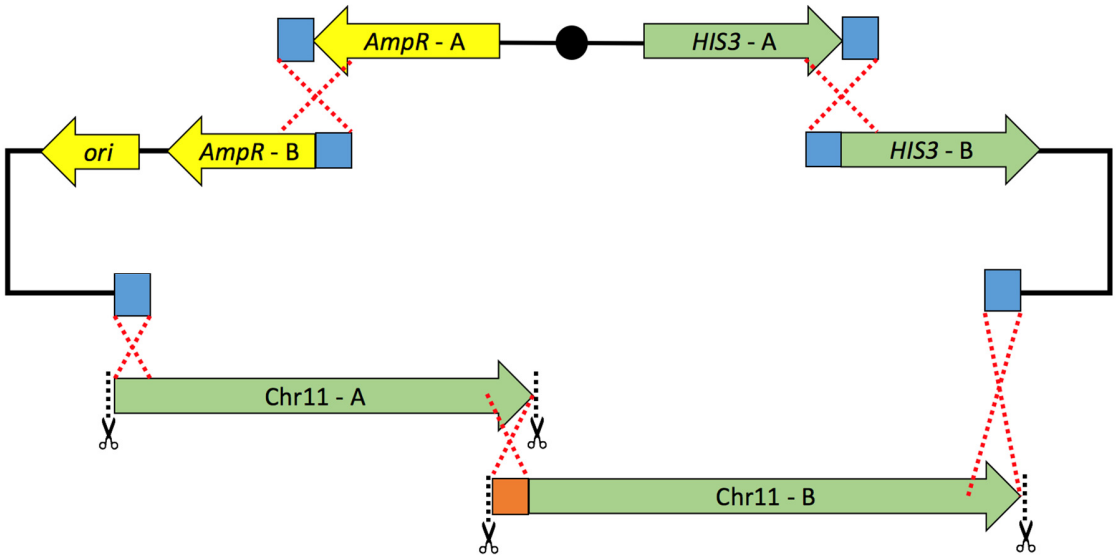


Figure 4.3: Schematic representation of the process used to combine two Chr11 tRNA array parts into a pRS413 backbone. The figure illustrates the overall process used to combine split tRNA arrays for the tRNA neochromosome (these tRNA arrays are summarised in **Table 4.1**). The pRS413 shuttle vector (upper portion of the diagram; 4,970 bp) was amplified using PCR with ~40 bp homology overhangs (blue boxes) for either Gibson Assembly or *in vivo* homologous recombination in BY4741. The black dot indicates the centromere and *ori* describes the origin of replication for propagation in *E. coli*. The 100 bp overlapping region for the Chr11 tRNA array (orange box) facilitated seamless assembly of the two tRNA array parts (5,136 bp and 5,201 bp, respectively), each of which were removed from their growth vectors using restriction digest (scissors with dotted lines).

Table 4.1: Summary of assembly strategies used to combine each split tRNA array 'part' and introduce them into pRS413. Yeast assembly refers to *in vivo* homologous recombination in *S. cerevisiae*. The Chr7 tRNA array, due to its large size, was designed to be split into three pieces.

tRNA Array	Size (Kb)	Assembly Strategy
Chr4	18.16	Yeast assembly
Chr5	12.91	Gibson Assembly
Chr7	23.28	Yeast assembly
Chr10	15.53	Yeast assembly
Chr11	10.34	Gibson Assembly
Chr12	13.63	Gibson Assembly
Chr13	13.56	Gibson Assembly
Chr15	12.90	Gibson Assembly
Chr16	10.97	Gibson Assembly

4.2.2. Neochromosome Construction Process Overview: Inchworming

The power of *in vivo* homologous recombination in *S. cerevisiae* to manipulate and join DNA sequences is well known, and is exploited by members of the Sc2.0 consortium to construct synthetic yeast chromosomes by swapping pre-existing sections of DNA with their synthetic counterpart (Annaluru *et al.*, 2014). An alternate strategy, also based on *in vivo* homologous recombination, was developed to construct the tRNA neochromosome from its constituent tRNA arrays.

This construction strategy was called *Inchworming* based on the characteristic locomotion of caterpillars of the Geometer moth, although it should be noted that a similar terminology exists called "Inchworm Elongation" that was used for genome construction of *Synechocystis* in *Bacillus subtilis* (Itaya *et al.*, 2005). Fundamentally, the inchworming method applied *in vivo* homologous recombination in *S. cerevisiae* to

incorporate new DNA sequences into a growing circular chromosome (essentially, a megaplasmid) prior to linearisation using the telomerator system (Mitchell and Boeke, 2014). To illustrate this process, **Figure 4.4** describes the method of combining the Chr6 (6.3 kb), Chr5 (13 kb) and Chr4 (18.2 kb) tRNA arrays to generate a ~45 kb construct.

Inchworming was designed to be induced by repeated swapping of the *LEU2* and *URA3* auxotrophic markers mediated by the presence of two homology arms. The first homologous region was the outermost (last) ~400 bp to ~600 bp DNA sequence of the tRNA array already present in the growing neochromosome. This region was designed to be dynamic, and alternate following every round of inchworming. In **Figure 4.4**, these regions are indicated by the blue t6.1 arrow (representing the outermost tRNA cassette of the Chr6 tRNA array) and the purple *Ter* replication termination site. The second homology region was a Universal Homologous Region (UHR), consisting of a 500 bp sequence of randomly-generated nucleotide bases. The UHR region (indicated by a red box in **Figure 4.4**) is static, and will not change after each round of transformation.

Following transformation of the 'inchworm fragment' into *S. cerevisiae*, these two homologous regions were designed to recombine, facilitating the swapping of either the *LEU2* or *URA3* selective markers and the integration a new piece of DNA into the growing circular neochromosome. The *LEU2* and *URA3* selective markers may be repeatedly recycled, allowing the repeated integration of any new piece of DNA. Finally, the static *HIS3* marker adjacent to the centromere provided an additional means of selection against re-circularisation events or genome integration.

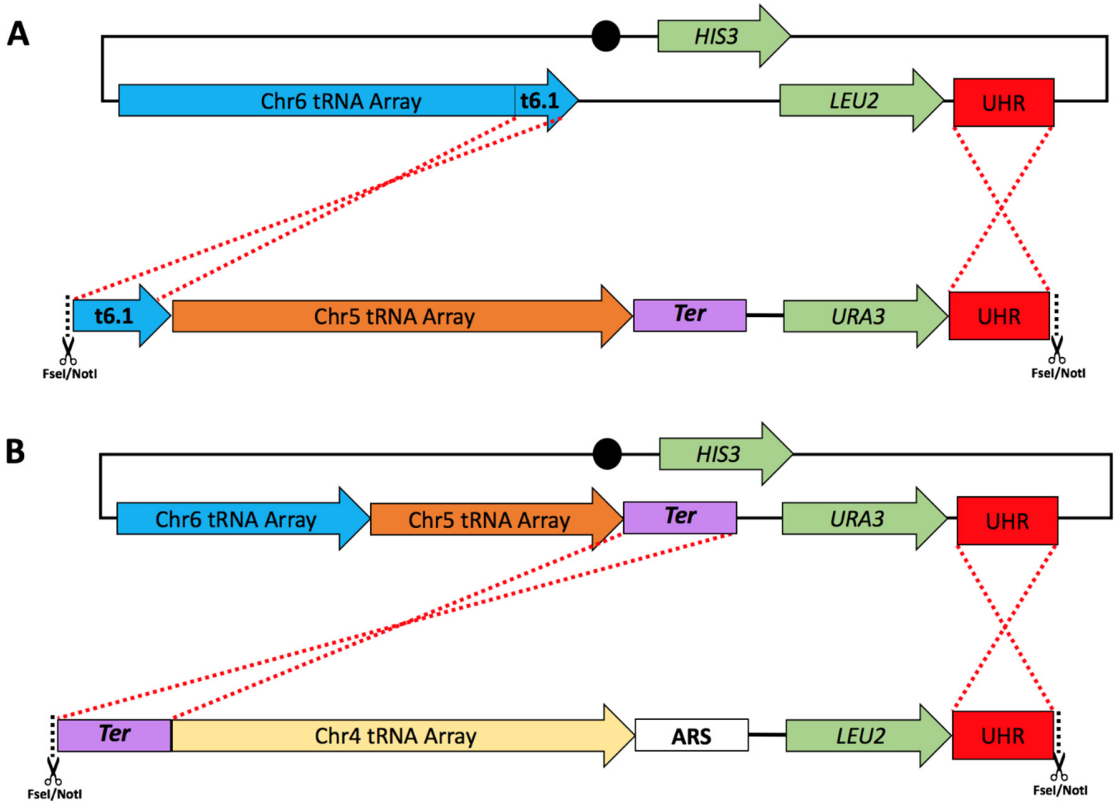


Figure 4.4: Schematic overview describing the first two rounds of neochromosome construction using the inchworming method. The figure illustrates the first two rounds of neochromosome construction using the inchworming method, describing the introduction of the Chr5 **(A)** and Chr4 **(B)** tRNA arrays into the growing neochromosome. The black dot indicates the centromere, tRNA arrays are indicated by coloured arrows, auxotrophic markers by green arrows and the 500 bp UHR with a red box. The FseI/NotI restriction sites are indicated by scissors and black dotted lines. (**A**) describes the first round of inchworming the Chr5 tRNA array (12,911 bp) into the neochromosome housing the Chr6 tRNA array (14,140 bp). The blue 't6.1' arrow (650 bp) is the last tRNA cassette of the Chr6 tRNA array. The blue 't6.1' arrow and the UHR sequence were designed to recombine, facilitating the swapping of the *LEU2* marker with the *URA3* marker. (**B**) Indicates the subsequent round of inchworming of the Chr4 tRNA array (18,160 bp). In this instance, the purple *Ter* box represents the 543 bp *Ter1* replication termination cassette and was used as one of the two homology arms. Following transformation, the *URA3* marker replaced the *LEU2* marker, facilitating the integration of the Chr4 tRNA array and generating a construct size of 45,425 bp. The white ARS box contains the *chrF-444* origin of replication (328 bp), and was the homologous region for the subsequent round of inchworming.

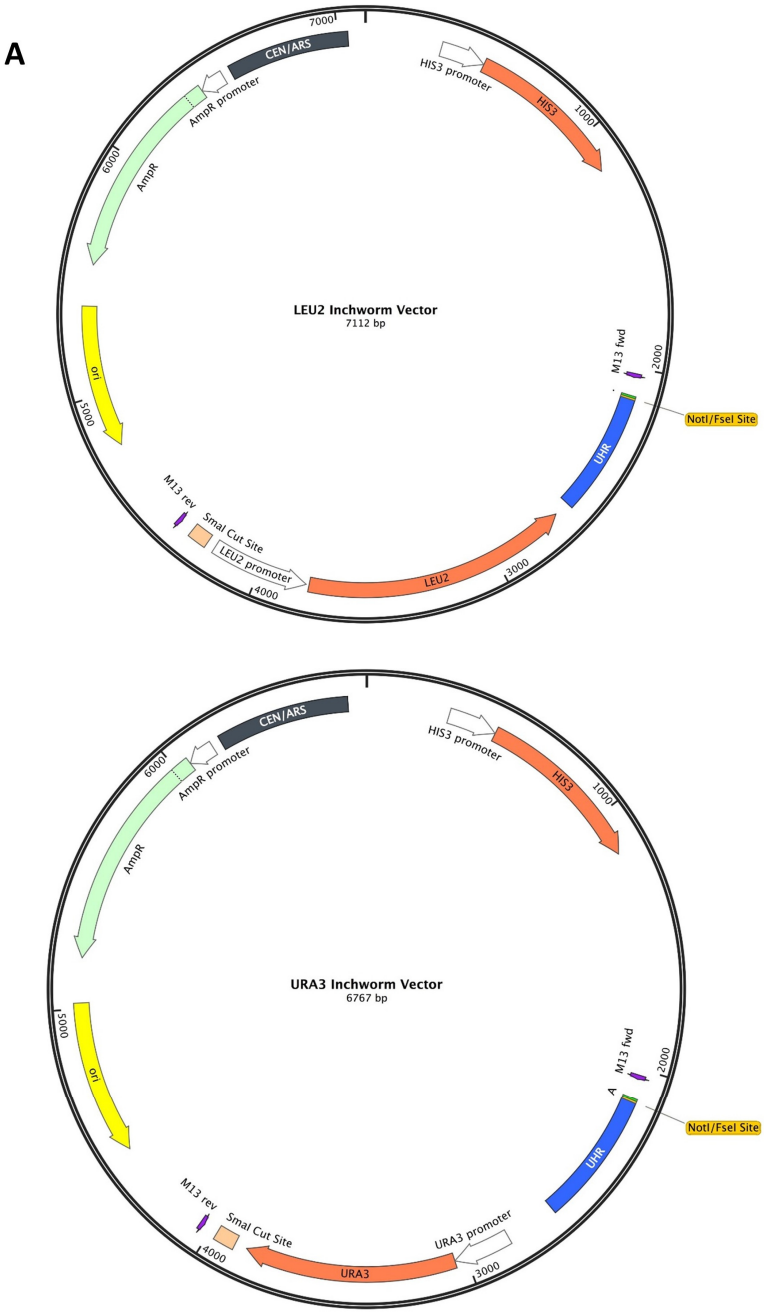
4.2.3. Approach Used to Construct Universal Inchworm Vectors

Fundamental to the process of inchworming were two "universal inchworm vectors" (**Figure 4.5A**) designed to facilitate the process of neochromosome construction. To produce sufficient DNA quantity for the inchworming process, each vector was designed to be amplified in *E. coli* so that the DNA sequences necessary for inchworming ('inchworm fragments') may simply be excised from their backbone using restriction digest prior to yeast transformation. Each vector was designed to house the genetic elements required for the inchworming process, including either a *LEU2* or *URA3* auxotrophic marker and a 500 bp UHR. The presence of a SmaI blunt-cutting restriction site facilitated the linearisation of each vector prior to allow the insertion of each tRNA array.

Section 4.3.2 briefly summarises the construction of each universal inchworm vector. To minimise any homology to the yeast genome, the UHR region was designed using a Perl script to produce a 500 bp sequence of randomly-generated DNA (**Figure 4.5B**) and subsequently constructed using the Build-a-Genome protocol. To verify a lack of homology to the *S. cerevisiae* genome, the UHR sequence was subjected to a BLAST search – only minimal homology was observed (data not shown). The *LEU2* and *URA3* inchworm vectors were then constructed by PCR-amplifying the pRS413 backbone, UHR and *LEU2/URA3* auxotrophic markers using primers designed to include overlapping 40 bp regions of homology. Following Gibson Assembly, inchworm vectors were verified using restriction digest DNA fingerprinting and Sanger sequencing across junction regions.

The presence of flanking FseI or NotI restriction sites was critical to the recovery of each inchworm fragment and were identified as they are rare cutters: if one enzyme was found to cut internally into the inchworm fragment, the second may be selected, and *vice versa*.

The first of these two restriction sites were generated by their inclusion in primers used to amplify the first bridge sequence; the second two restriction sites were already present in the inchworm vector.



[Figure continued on following page]

B

```perl
#!/usr/bin/perl
# Roys Random DNA Generator
use strict;
use warnings;

srand(time|$$);

my @random_DNA = ();
my($number_of_sequences) = 10;
my $dna;

for (my $i = 0; $i < $number_of_sequences ; ++$i) {
    my($length) = 500;
    my $dna;
        for (my $i=0 ; $i < $length ; ++$i) {
        my(@nucleotides) = ('A', 'T', 'G', 'C');
        $dna .= $nucleotides [rand @nucleotides];
    }
    print "$dna\n\n";
}
```

Figure 4.5: Plasmid maps of the *LEU2* and *URA3* universal inchworm vectors used to facilitate neochromosome construction and Perl script used to generate the UHR sequence. (**A**) Each inchworm vector was based on the pRS413 shuttle vector and contained the genetic elements required for the inchworming process. Each inchworm vector is 7,112 bp and 6767bp, respectively. The *LEU2*, *HIS3* and *URA3* auxotrophic markers are indicated by orange arrows, the UHR region by a blue box, SmaI cut sites by a beige box, CEN/ARS by a black box, bacterial origin (*ori*) by a yellow arrow, bacterial ampicillin resistance gene (*AmpR*) indicated by a green arrow and NotI/FseI restriction sites by an orange box. The M13 Fwd and Rev primer binding sites were used to verify the above constructs with Sanger sequencing. (**B**) Simple Perl programming script used to produce a 500 bp stretch of randomly-generated DNA. The script utilised a random number generator to produce a series of ten, 500 bp DNA sequences containing random nucleotides.

4.2.4. Strategy Used to Introduce tRNA Arrays into Inchworm Vectors

Prior to inchworming, it was necessary to introduce each tRNA array into an inchworm vector. To do so, each tRNA array was first recovered from their pRS413 backbone using restriction digest and combined with either the *LEU2* or *URA3* inchworm vector linearised with the SmaI restriction enzyme. Assembly of these two components was mediated by two 330 bp to 650 bp 'bridge' sequences designed to act as regions of homology for inchworming. As summarised in Section 4.2.2, the first bridge sequence was the last segment of DNA from the latter round of inchworming, and the second bridge sequence

was the first segment of DNA necessary for the subsequent round of inchworming. **Figure 4.6** summarises the primary steps involved in introducing tRNA arrays into their respective inchworm vectors. **Figure 4.7** provides a graphical representation of this process and Section 4.3.3 describes a representative example used to introduce the Chr10 tRNA array into the *LEU2* inchworm vector.

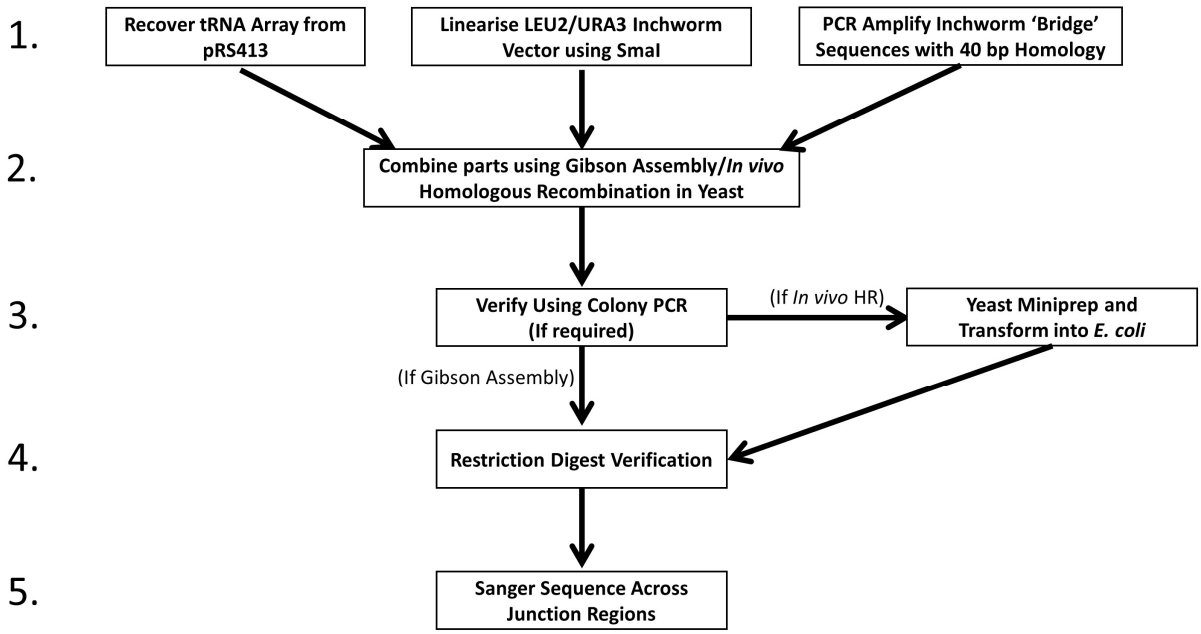


Figure 4.6: Flowchart summarising the primary steps involved in introducing tRNA arrays into their respective inchworm vectors. As described in step 3 on the above diagram, either Gibson Assembly or *in vivo* homologous recombination in yeast (*in vivo* HR) may be used to generate each construct.

Each bridge sequence was generated by PCR-amplifying the necessary sequence with primers designed to house two 40 bp overhangs (the 'A' and 'B' boxes in **Figure 4.7**). These 40 bp overhangs provided the means to combine each tRNA array with either the *URA3* or *LEU2* inchworm vector using a combination of either Gibson Assembly or *in vivo* homologous recombination in BY4741. These bridge sequences varied depending on the homology region required for inchworming, and included either tRNA cassettes, ARS

elements or *Ter* cassettes (**Figure 4.7** illustrates the use of ARS and *Ter* cassettes used to facilitate inchworming).

Following integration of each tRNA array, the junction regions of integration were verified using Sanger sequencing. **Table 4.2** summarises the assembly methods used to introduce each tRNA array into inchworm vectors.

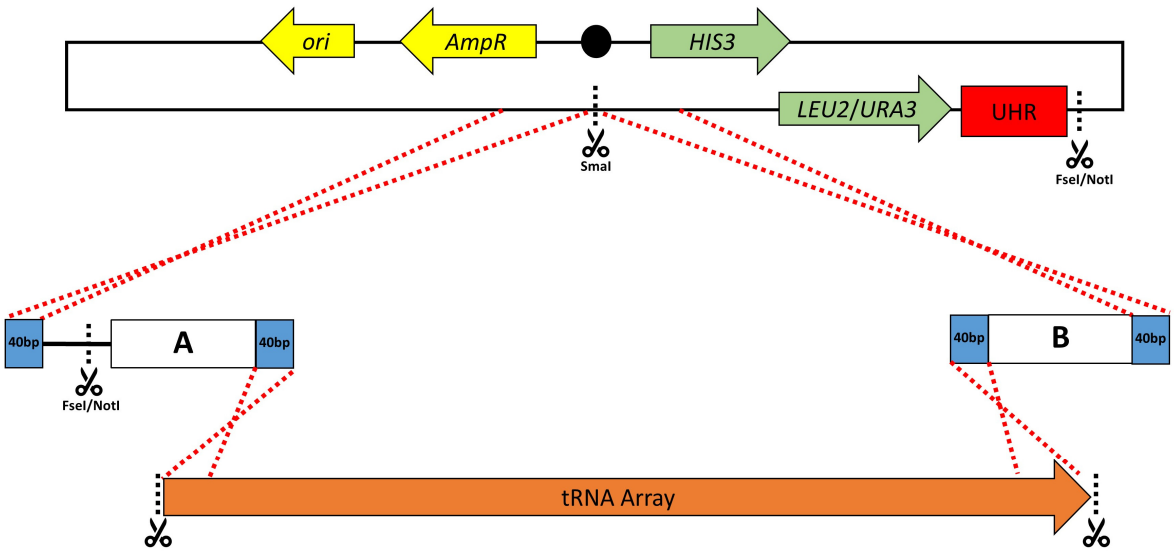


Figure 4.7: Schematic representation of the approach used to insert tRNA arrays into the *LEU2* or *URA3* inchworm vectors. The figure describes the general strategy using either Gibson assembly or *in vivo* homologous recombination in yeast to introduce tRNA arrays into either the *LEU2* or *URA3* inchworm vectors, facilitated by two 'bridge' sequences. The white A and B boxes represent the homologous bridge sequences used for each round of inchworming and are amplified using PCR with ~40 bp homology overhangs to both the tRNA array and the *LEU2*/*URA3* inchworm vector. Scissors with dotted lines represent restriction enzyme cut sites. Following linearisation of the inchworm vector using the SmaI restriction enzyme, the presence of ~40 bp homology arms (blue boxes) on the inchworm bridge sequences (white A and B boxes) facilitated the integration of tRNA arrays (orange arrow) using a combination of either Gibson Assembly or *in vivo* homologous recombination in BY4741. The presence of a FseI or NotI restriction sites allowed the recovery of the entire inchworm fragment prior to yeast transformation.

Table 4.2: Order of neochromosome construction and inchworm vector used for each tRNA array. The table summarises the order of neochromosome construction, and provides ancillary information on each tRNA array including the assembly strategy used to introduce each into their respective inchworm vectors. 'Order of inchworming' in the table below refers to the order of integration used to construct the neochromosome, with the Chr6 tRNA array representing the start point of construction. 'Size' refers to the size of each tRNA array, 'cutters' refers to the flanking restriction sites used to retrieve each DNA fragment for each round of inchworming. 'Assembly strategy' refers to the approach used to introduce each tRNA array into their respective inchworm vector, and includes, Gibson Assembly, Yeast assembly (*in vivo* homologous recombination in BY4741) or ligation for the Chr6 tRNA array.

Order of Inchworming	tRNA Array	Size (Kb)	Inchworm Vector	Cutters	Assembly Strategy
1	Chr6	6.49	*LEU2*	N/A	Ligation
2	Chr5	12.91	*URA3*	FseI	Gibson
3	Chr4	18.16	*LEU2*	NotI	Yeast
4	Chr3	6.46	*URA3*	FseI	Gibson
5	Chr2	8.41	*LEU2*	NotI	Gibson
6	Chr1	2.63	*URA3*	NotI	Gibson
7	Chr7	23.28	*LEU2*	NotI	Yeast
8	Chr16	10.97	*URA3*	NotI	Gibson
9	Chr9	6.46	*LEU2*	NotI	Yeast
10	Chr8	7.13	*URA3*	NotI	Gibson
11	Chr10	15.53	*LEU2*	FseI	Yeast
12	Chr11	10.34	*URA3*	NotI	Gibson
13	Chr13	13.56	*LEU2*	NotI	Yeast
14	Chr12	13.63	*URA3*	NotI	Yeast
15	Chr14	9.08	*LEU2*	NotI	Gibson
16	Chr15	12.90	*URA3*	NotI	Gibson

4.2.5. Neochromosome Construction and Verification through PCR Tag Analysis

To initiate construction of the neochromosome, it was necessary to include a suitable 'start point' for neochromosome construction. The *LEU2* inchworm vector, based on a pRS413 backbone, was selected due to the presence of a *LEU2* marker, centromere and UHR region necessary for inchworming. The Chr6 tRNA array was introduced into the *LEU2* inchworm vector *via* ligation and represented the start-point for construction of the neochromosome (data not shown). Following ligation, this construct was verified using restriction digest DNA fingerprinting and Sanger sequencing.

It was decided at an early stage to construct the neochromosome in two separate *S. cerevisiae* strain backgrounds. Both BY4741 (RWy007) wild-type and a strain housing a triple-synthetic chromosome III, VI and right arm of IX (RWy121) were used to construct two independent versions of the neochromosome in parallel. It was postulated that the lack of tRNA genes in the triple-synthetic-chromosome strain may aid stability of the neochromosome during construction.

Neochromosome construction was performed in a stepwise manner utilising repeated rounds of *LEU2* and *URA3* marker swapping in yeast. Each tRNA array and sequences necessary for inchworming were first recovered from their backbone using restriction digest using either NotI or FseI, gel purified, and transformed into yeast containing the growing circular neochromosome. Transformants were plated onto the required selective media and incubated at 30^{o}C until the appearance of sufficiently large colonies. The order of each round of integration is shown in **Table 4.2**.

To verify the presence of each tRNA gene after each round of inchworming, a series of primers ("PCR Tags") were manually designed using Snapgene software (Chicago, USA) to anneal with the 500 bp 5' and 40 bp 3' region of each tRNA cassette to produce an

amplicon of approximately 400 bp to 600 bp (**Figure 4.8**). After each round of inchworming, a full set of PCR tags were utilised to verify the overall integrity of the neochromosome. The full list of PCR tags may be found in Appendix III, **Table II**.

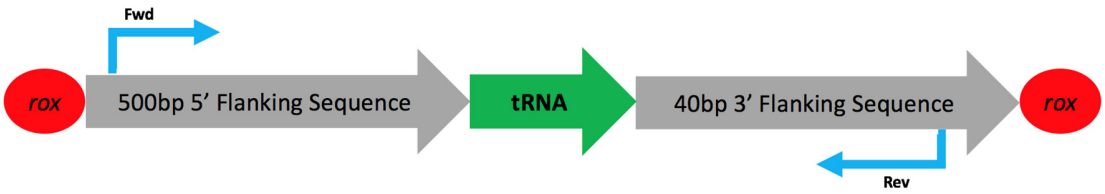


Figure 4.8: Representative diagram indicating the method used to design PCR tag primer pairs. The first primer was designed to anneal with the 5' flanking sequence (blue arrow on the left), the second primer (blue arrow on the right) was designed to anneal with the 3' flanking sequence of each tRNA gene, producing an amplicon of approximately 400 bp to 600 bp. A full list of PCR tags used may be found in Appendix III, **Table II**.

Finally, as the neochromosome gradually increased in size, a workflow was developed to verify the integration of each tRNA array. This process included the induced loss of an unwanted secondary selective marker following the previous round of inchworming, as well as multiple rounds of screening using colony PCR. This is illustrated in **Figure 4.9**.

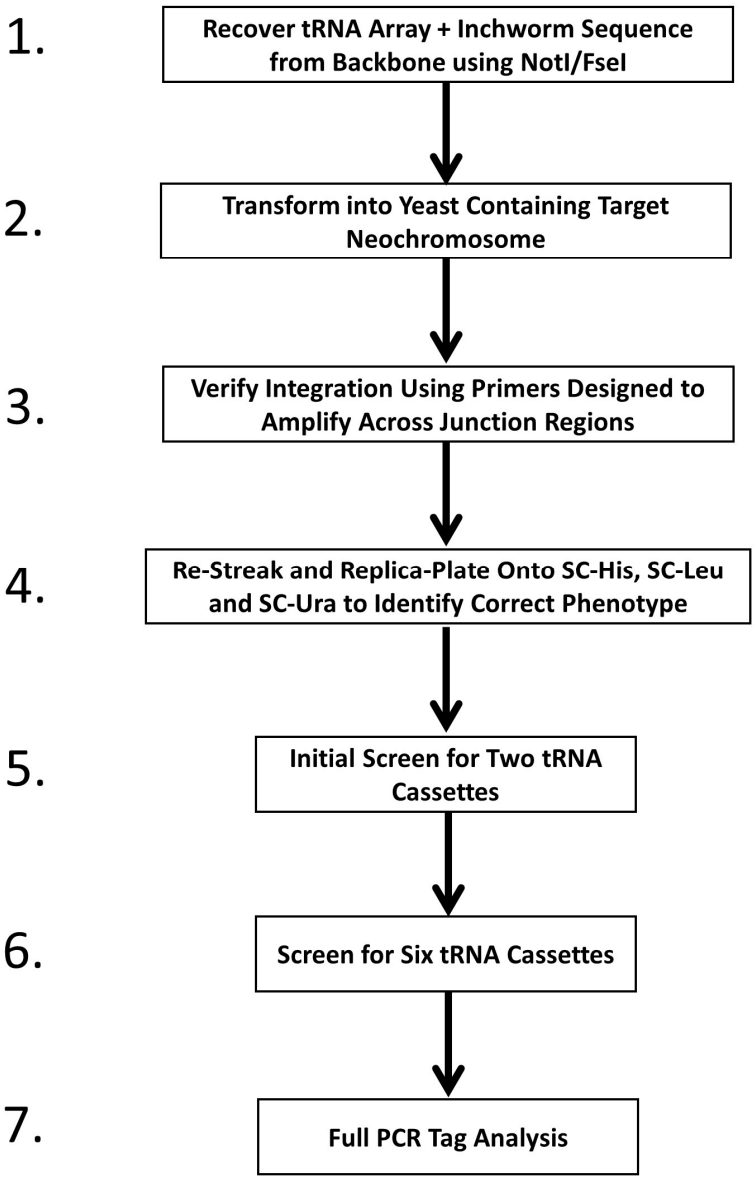


Figure 4.9: Flowchart summarising the primary steps involved in the construction of the tRNA neochromosome using the inchworming method. The tRNA array in its inchworm vector in step 1 is designed to house homology arms and the necessary selective marker (*LEU2* or *URA3*) to swap with the existing marker in the growing tRNA neochromosome (step 2).

4.3. Results

4.3.1. Combining the Split Chr11 tRNA Array and Introduction into pRS413

Results demonstrating the construction of the Chr11 tRNA array from its constituent halves is shown in **Figure 4.11** as a representative example of the strategy used for the other tRNA arrays. A schematic describing the approach used below may be found in **Figure 4.3**. The pRS413 backbone was first split into three fragments using PCR (**Figure 4.11A**), with primers designed to house homology to the other fragments and two tRNA array halves (Illustrated in **Figure 4.10** and listed in **Table 4.3**). The two halves of the Chr11 tRNA array (5,136 bp and 5,201 bp, respectively) were then recovered using restriction digest with NotI (**Figure 4.11B**) before combining each gel-purified fragment using Gibson Assembly. Following Gibson Assembly, bacterial colonies were subjected to restriction digest verification using SacI (**Figure 4.11C**) and Sanger sequencing across junction regions (**Figure 4.11D**). Colony PCR to verify the presence of the Chr11 tRNA array was not performed in this instance. Primers used for Sanger sequencing (M13F, M13R and YCp3084) are denoted in **Figure 4.10**. The results show the successful construction of the Chr11 tRNA array from its two halves.

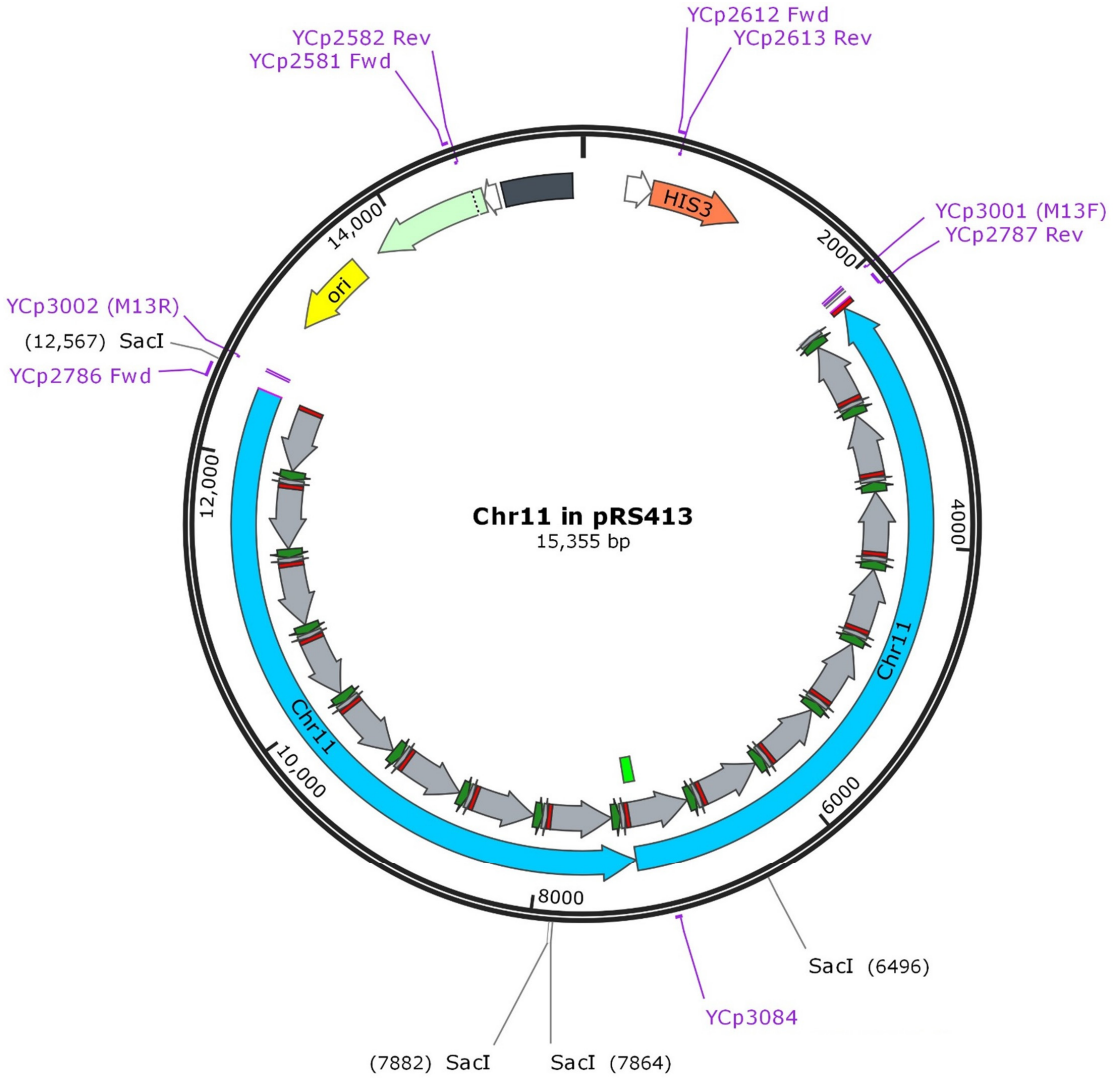


Figure 4.10: Plasmid map of the Chr11 tRNA Array in pRS413. This complete tRNA array is 15,335 bp and was synthesised in two sections (blue arrows; 5,136 bp and 5,201 bp, respectively), and combined using Gibson Assembly. Primers on the above diagram are indicated by pink lines, and restriction sites used to verify the above construct using SacI (**Figure 4.11C**) are denoted by black lines (the coordinates of the restriction cut sites are in brackets). The black box on the above diagram represents the CEN/ARS sequence, the teal arrow denotes the *AmpR* marker, the orange arrow shows the *HIS3* marker and the yellow arrow is the *ori*. The central overlapping region to combine the two tRNA halves is indicated in green, with primers used to amplify the pRS413 backbone shown in **Table 4.3**. tRNA cassettes are indicated by grey arrows (tRNA flanking sequences), green arrows (tRNA genes) and red blocks (*rox* recombination sites). Flanking the tRNA array sequence lie NotI restriction sites (not shown), facilitating removal of the tRNA array from the pRS413 backbone.

Table 4.3: PCR primers used to amplify the pRS413 backbone prior to Gibson assembly with the two halves of the Chr11 tRNA array. The 'Reaction' column in the table below refers to wells 1, 2 and 3 in **Figure 11A** and are illustrated on **Figure 4.10**. The sequences for each primer may be found in **Appendix VI**.

Reaction	Forward Primer	Reverse Primer	Expected Amplicon Size (bp)
pRS413-1	YCp2786	YCp2582	2,100
pRS413-2	YCp2581	YCp2613	1,502
pRS413-3	YCp2612	YCp2787	1,574

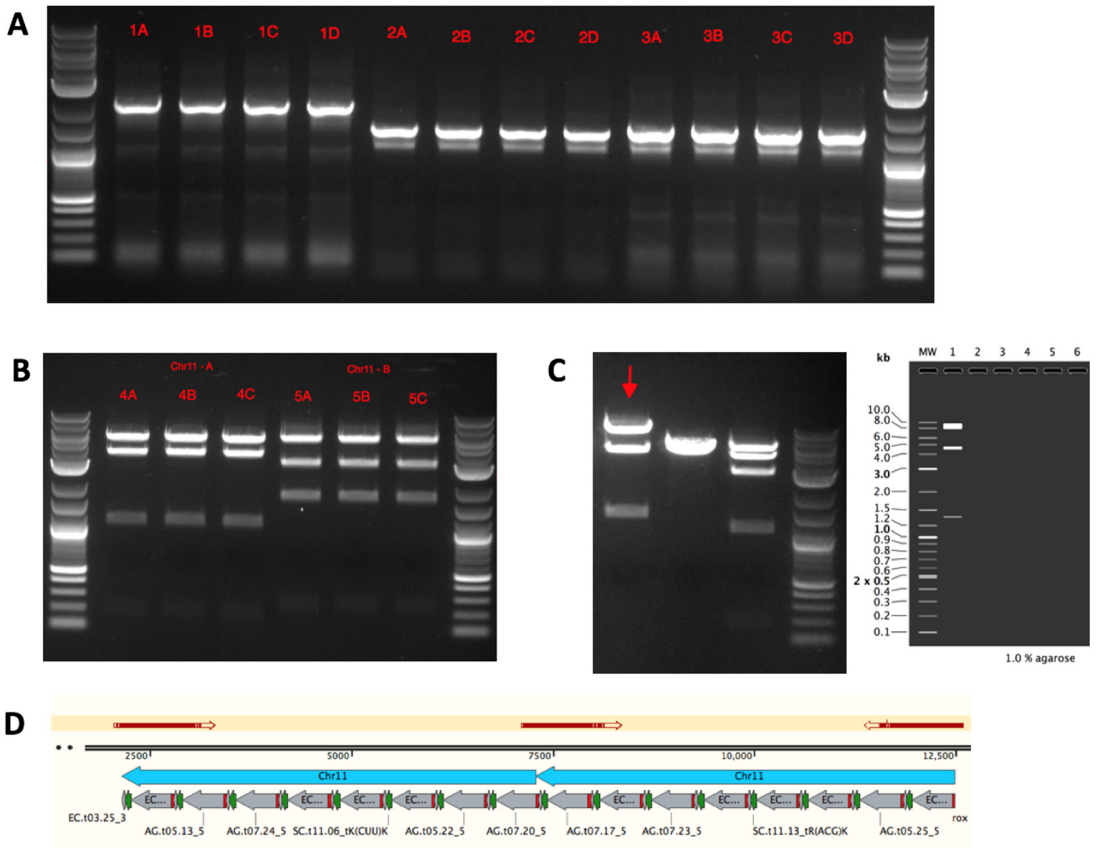


Figure 4.11: Representative results describing the process used to combine two Chr11 tRNA array halves into a pRS413 backbone. The figure shows the individual steps used to successfully combine the two halves of the Chr11 tRNA array using Gibson Assembly. PCR reactions and restriction digests for **(A)** and **(B)** were undertaken as replicates and later pooled together to maximise the quantity of DNA recovered following gel extraction. A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1**.

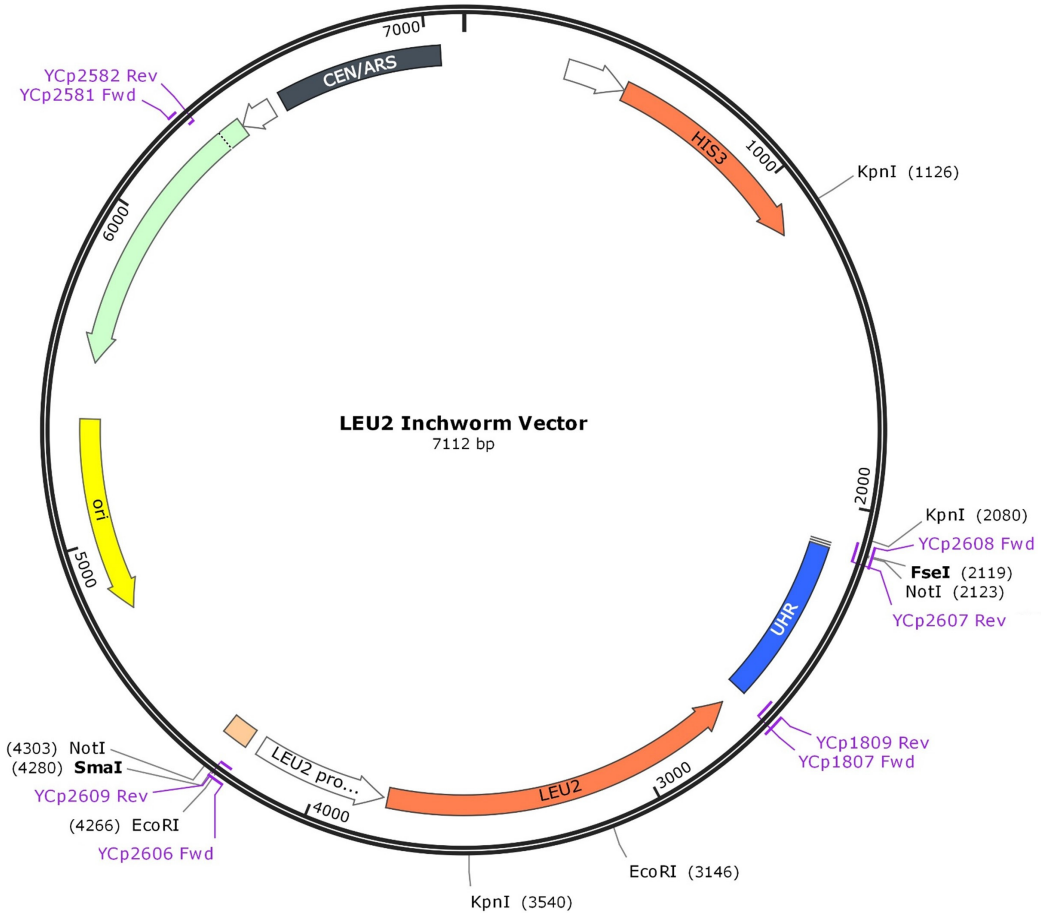
[Figure legend continued on Following Page]

(Figure 4.11 continued) (**A**) PCR Amplification of pRS413 into three individual parts (lanes 1, 2 and 3; n = 4), with each fragment designed to house overhangs for either Gibson Assembly or *in vivo* homologous recombination in yeast. The location of primer binding sites is illustrated in **Figure 4.10**, and listed in **Table 4.3** with expected amplicon sizes. Nonspecific amplification was observed for parts 2 and 3; the upper bands were carefully excised, taking care to avoid the lower bands. (**B**) Recovery of the two Chr11 tRNA array halves from their growth vector (lanes 4 and 5; n = 3). As the two tRNA array halves (5,136 bp and 5,201 bp, respectively) were found to be of similar size to their pRS413 backbone (4,970 bp), secondary restriction enzymes in addition to NotI (HindIII for part A and PvuI for part B) were selected to internally cut each pRS413 backbone into smaller fragments. The expected restriction digest fragment sizes are therefore: 5,136 bp, 3,615 bp, 1,117 bp, 187 bp and 49 bp (part A) and 5,201 bp, 3,054 bp, 1,678 bp and 238 bp (part B). The upper bands were gel excised. (**C**) Restriction Digest verification of the Chr11 tRNA array following Gibson Assembly (n = 3). The SacI enzyme was used to cut both Chr11 tRNA array halves internally (the location of each cut site may be found in **Figure 4.10**), providing a reliable indicator of successful construction. The expected restriction digest fragment sizes are 9,284 bp, 4,685 bp, 1,368 bp and 18 bp. The red arrow indicates the correct restriction digest pattern and the predicted restriction digest is shown to the right – slight differences in the real and predicted upper band sizes may be potentially caused by differing concentration of DNA loaded onto the agarose gel (deep sequencing following neochromosome construction reveals no alterations to the Chr11 tRNA array). (**D**) Representative diagram following Sanger sequencing verification. Primers were designed to produce sequencing reads across junction regions (illustrated in **Figure 4.10**). Sequencing regions (red arrows) were visually checked for base errors.

4.3.2. Constructing Universal Inchworm Vectors

The *LEU2* and *URA3* inchworm vectors were constructed by PCR amplifying pRS413, *LEU2* and *URA3* markers and the UHR before combining each using Gibson Assembly. The primers used to amplify each fragment are illustrated in **Figure 4.12** and listed **Table 4.4**. **Figure 4.13** describes the verification of each construct, including the assembly of the 500 bp UHR from its constituent oligonucleotides using the Build-a-Genome protocol (**Figure 4.13A**). The successfully amplified UHR region can be observed as strong bands around the 500 bp mark on the gel. Following Gibson Assembly, four bacterial colonies of each were subjected to restriction digest DNA fingerprinting (**Figure 4.13B**) and Sanger sequencing (**Figure 4.13C**) using the M13 forward and reverse primers. The restriction

enzyme cut sites used to verify each construct are illustrated in **Figure 4.12**. The results show the successful construction of the *LEU2* and *URA3* inchworm vectors.



[Figure continued on following page]

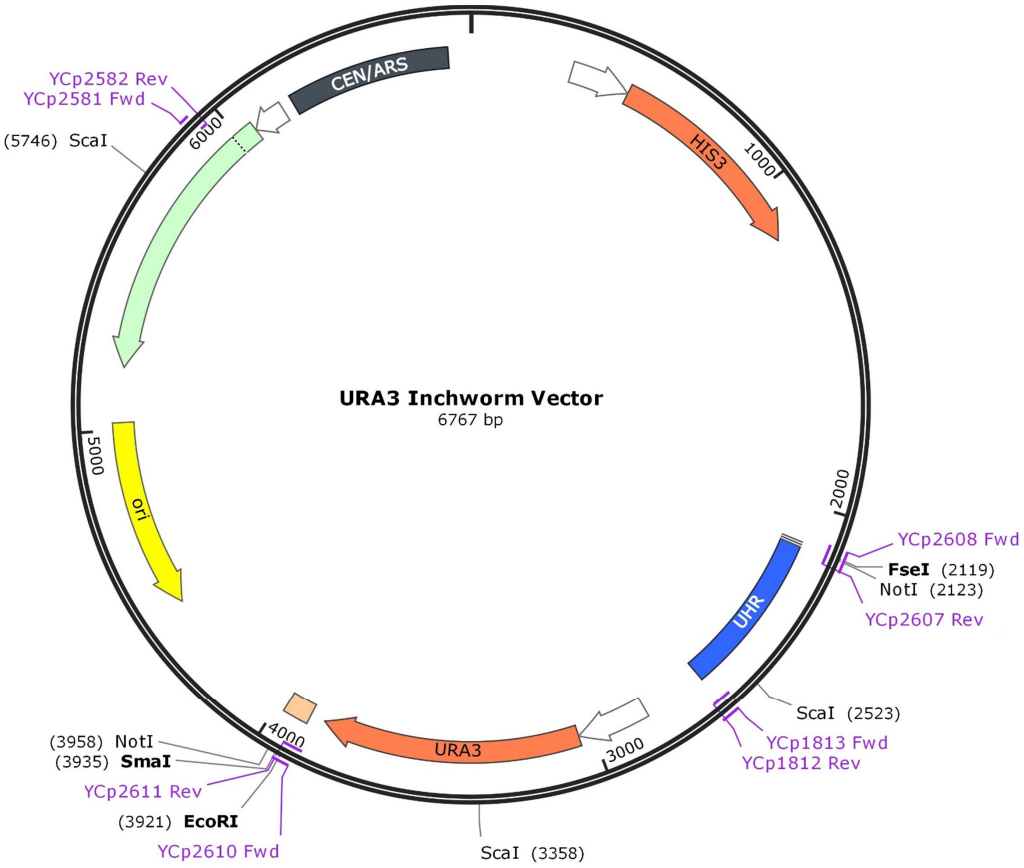


Figure 4.12: Plasmid maps of the *LEU2* and *URA3* inchworm vectors. Each vector was based on the pRS413 shuttle vector and contains the genetic elements required for the inchworming process. The *LEU2*, *HIS3* and *URA3* auxotrophic markers are indicated by orange arrows, the UHR region by a blue box, SmaI cut sites by a beige box, CEN/ARS by a black box, bacterial origin (*ori*) by a yellow arrow and bacterial ampicillin resistance gene (*AmpR*) indicated by a teal arrow. All restriction enzyme sites used to verify the above constructs (**Figure 4.13B**), insert inchworm vectors (SmaI) and recover the inchworm fragments prior to inchworming (FseI/NotI) are indicated by black arrows (the coordinates of the restriction cut sites are in brackets). Primers used to generate the above construct are denoted with pink arrows and are summarised in **Table 4.4**. M13 Fwd and Rev primes were used to verify the above constructs with Sanger sequencing (**Figure 4.13C**).

Table 4.4: PCR primers used to construct the *LEU2* and *URA3* inchworm vectors prior to Gibson assembly. *LEU2* and *URA3* in the 'Reaction' column below refer to the fragments of the *LEU2* and URA3 inchworm vectors, respectively. Each fragment may be found on **Figure 4.12**. The sequences for each primer may be found in **Appendix VI**.

Reaction	Forward Primer	Reverse Primer	Template	Expected Amplicon Size (bp)
LEU2-1	YCp2606	YCp2582	pRS413	2,064
LEU2-2	YCp2581	YCp2607	pRS413	2,986
LEU2-3	YCp1809	YCp2609	pRS415	1,666
LEU2-4	YCp2608	YCp1807	UHR	556
URA3-1	YCp2610	YCp2582	pRS413	2,064
URA3-2	YCp2581	YCp2607	pRS413	3,007
URA3-3	YCp1813	YCp2611	pRS416	1,325
URA3-3	YCp2608	YCp1812	UHR	557

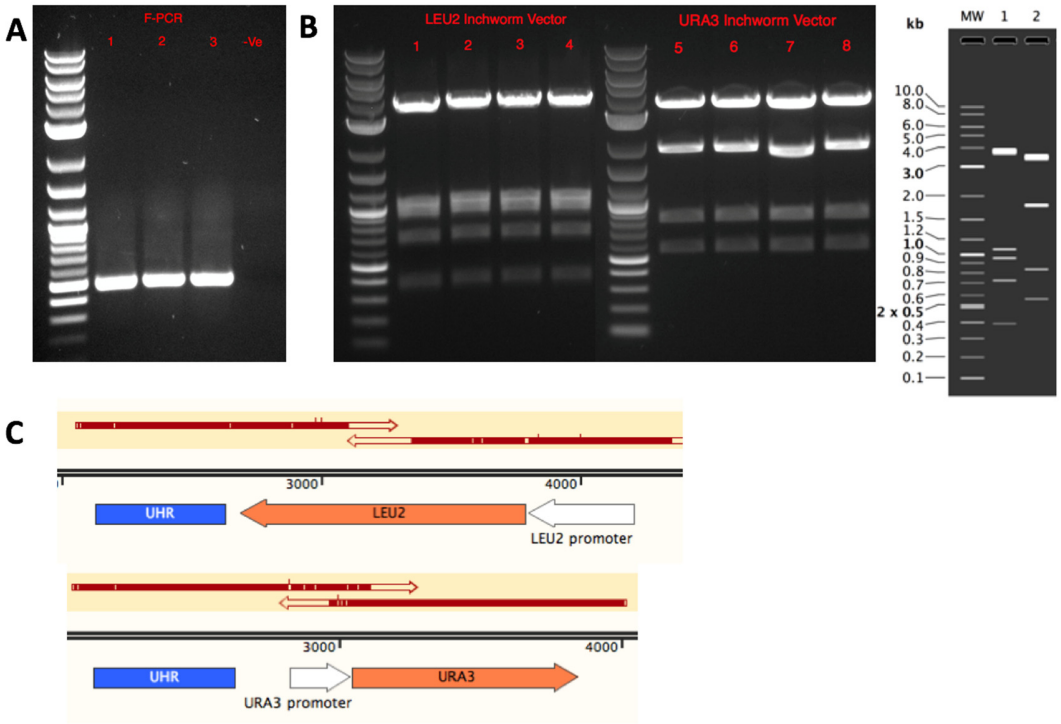


Figure 4.13: Verification of the *LEU2* and *URA3* inchworm vectors following Gibson Assembly. The above figure illustrates the construction of the UHR region and the verification of the two universal inchworm vectors. A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1**.

[Figure legend continued on following page]

Figure 4.13 (continued) (**A**) Finishing PCR (F-PCR) product displaying successful construction of the UHR from constituent oligonucleotides (n = 3). The 500 bp band representing the UHR is shown for F-PCR wells 1, 2 and 3 (**B**) Restriction-digest verification of inchworm vectors following Gibson Assembly (n = 4). KpnI + EcoRI and ScaI + EcoRI double-digests were used to produce the corresponding DNA fingerprint. The predicted restriction digest and fragment sizes of the 2-log ladder is shown to the right – lanes 1 and 2 are that of the *LEU2* and *URA3* inchworm vectors, respectively. The expected restriction digest fragment sizes for the *LEU2* inchworm vector are 3,972 bp, 1,066 bp, 954 bp, 726 bp and 394 bp, the expected restriction digest fragment sizes for the *URA3* inchworm vector are 3544 bp, 1,825 bp, 835 bp and 564 bp. The location of each restriction enzyme cut site is illustrated on **Figure 4.12** (**C**) Sanger sequencing of two isolates housing the *LEU2* inchworm vector (upper image) and the *URA3* inchworm vector (lower image) using M13 primers. Sanger sequencing runs are shown as red arrows. Sanger sequencing revealed several mutations or mismatches with the plasmid map, however, these did not affect functionality of markers or the UHR.

4.3.3. Introducing the Chr10 tRNA Array into the LEU2 Inchworm Vector

The introduction of the Chr10 tRNA array (15.5 kb) into the *LEU2* inchworm vector is shown in **Figure 4.15** as a representative example of the process undertaken for the other tRNA arrays. A schematic representation illustrating this process may also be found in **Figure 4.7**. The Chr10 tRNA array was first recovered from its pRS413 backbone using NotI (**Figure 4.15A**), before linearising the *LEU2* inchworm vector with SmaI (**Figure 4.15B**) and recovering each from an agarose gel using gel purification. The two inchworm bridge sequences were amplified using PCR with primers designed to contain the necessary homology arms for Gibson Assembly and subsequently gel-purified (**Figure 4.15C**). The primers used to amplify each inchworm bridge sequence are illustrated in **Figure 4.14** and listed in **Table 4.5**.

As Gibson Assembly failed to produce the expected construct, the Chr10 tRNA array inchworm vector was instead assembled using *in vivo* homologous recombination in BY4741. The tRNA array and *LEU2* inchworm vector were combined in equimolar

quantities, with the two inchworm bridge sequences added in excess prior to yeast transformation. Individual yeast colonies were carefully picked and tested for the presence of the Chr10 tRNA array using colony PCR screening for the SC.t10.07 tRNA cassette (**Figure 4.15D**), with four candidate isolates subjected to yeast plasmid extraction and subsequent restriction digest DNA fingerprinting using SpeI (**Figure 4.15E**). The location of these restriction cut sites is illustrated on **Figure 4.14**. Of the four isolates identified, three displayed the correct restriction digest pattern (wells 1, 2 and 4). The isolate corresponding to well 1 was then subjected to Sanger sequencing using the M13R (YCp3002) and YCp3100 primers across the two inchworm bridge sequences to verify successful construction and a lack of nucleotide variations in the bridge sequences (**Figure 4.15F**). The results show successful integration of the 15.53 kb Chr10 tRNA array into the *LEU2* inchworm vector.

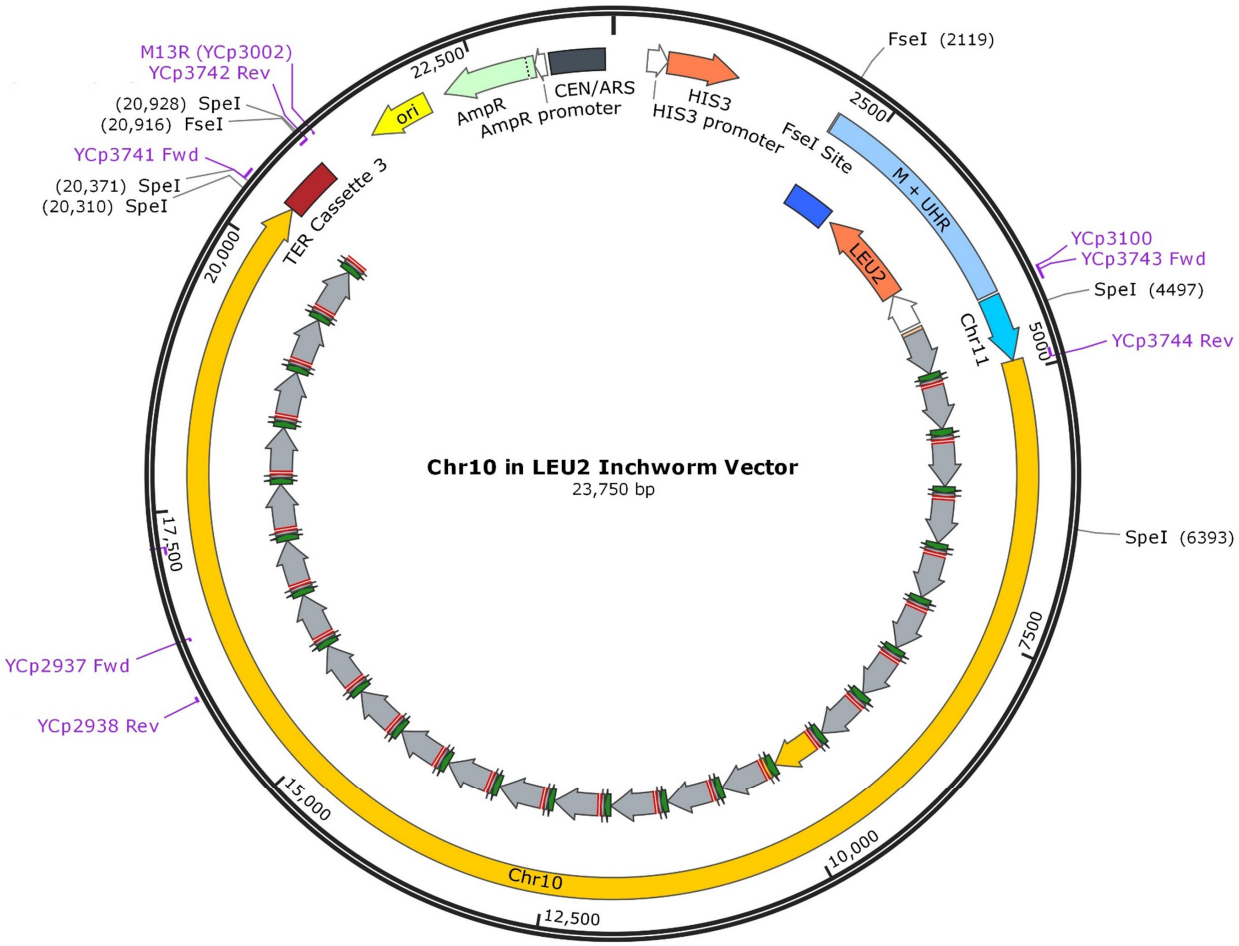
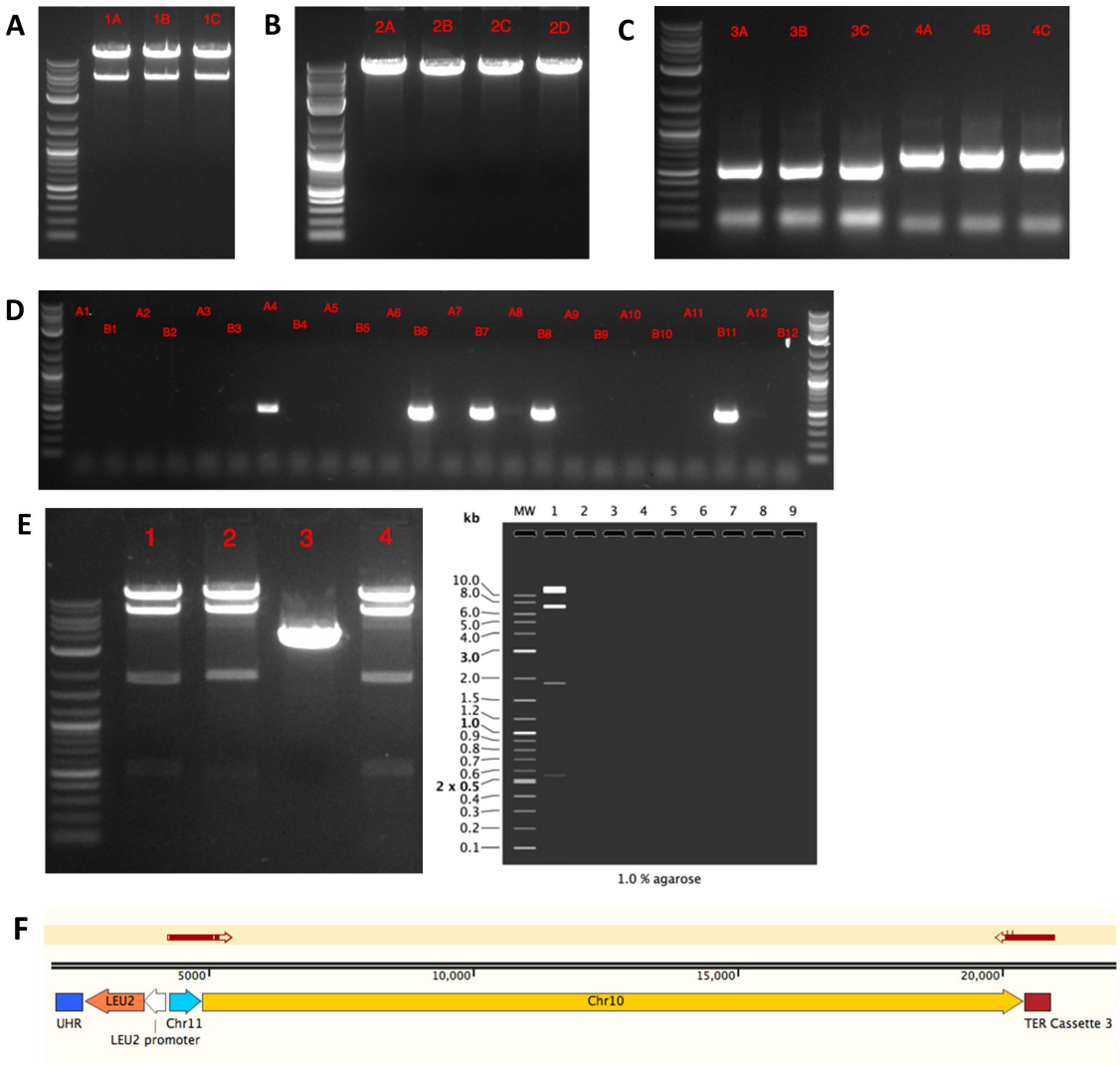


Figure 4.14: Plasmid map of the Chr10 tRNA Array in the *LEU2* inchworm vector. The above vector (23,750 bp) contains the Chr10 tRNA array and the genetic elements required for the inchworming process. The *LEU2* and *HIS3* auxotrophic markers are indicated by orange arrows, the UHR region by a blue box, CEN/ARS by a black box, bacterial origin (*ori*) by a yellow arrow and bacterial ampicillin resistance gene (*AmpR*) by a teal arrow. Primers on the above diagram are indicated by pink lines, and restriction sites used to verify the above construct using SpeI (**Figure 4.15E**) are denoted by black lines (the coordinates of restriction cut sites are in brackets). The FseI restriction cut sites used to excise the above inchworm sequence are also noted on the above plasmid map. The primers used to amplify the two 'bridge' sequences (*TER* Cassette 3 and the Chr11 tRNA array) are summarised in **Table 4.5**. M13R (YCp3002) and YCp3100 were used for Sanger sequencing across the inchworm bridge sequences (**Figure 4.15F)**. YCp2937 and YCp2938 were used for colony PCR to verify the presence of the Chr10 tRNA array (**Figure 4.15D).** All primer sequences may be found in **Appendix VI.**

Table 4.5: PCR primers used to amplify the inchworm 'bridge' sequences for the Chr10 inchworm vector. The following primers were used to amplify two 'bridge' sequences containing 40bp homology arms to facilitate introduction of the Chr10 tRNA array into the *LEU2* inchworm vector. 'Left' and 'Right' in the 'Reaction' column below refer to amplification of *TER* Cassette 3 and the SC.t11.01 tRNA cassette of the Chr11 tRNA array, respectively (**Figure 4.15C**). The sequences for each primer may be found in **Appendix VI**.

Reaction	Forward Primer	Reverse Primer	Template	Expected Amplicon Size (bp)
Left	YCp3741	YCp3742	*TER* Cassette 3	577
Right	YCp3743	YCp3744	Chr11 tRNA array	701



[Figure legend on following page]

Figure 4.15: Representative results describing the process of introducing the Chr10 tRNA array into the *LEU2* inchworm vector. The figure shows the individual steps used to successfully introduce the 15,540 bp Chr10 tRNA array into the 7,112 bp *LEU2* inchworm vector using *in vivo* homologous recombination in yeast. PCR reactions and restriction digests in **(A)**, **(B)** and **(C)** were undertaken as replicates and later pooled together to maximise the quantity of DNA recovered following gel extraction. A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1.** (**A**) Recovery of the Chr10 tRNA array from its pRS413 backbone using FseI (n = 3). Expected restriction digest fragment sizes are 15,540 bp and 4,978 bp. The upper bands corresponding to the Chr10 tRNA array were gel excised. (**B**) Linearisation of the *LEU2* inchworm vector using the SmaI restriction enzyme (n = 4). The expected restriction digest fragment size is 7,112 bp. (**C**) PCR amplification of bridge sequences containing 40 bp homology arms to facilitate introduction of the Chr10 tRNA array into the *LEU2* inchworm vector (n = 3). Wells 2A, 2B and 2C are reactions amplifying *Ter* cassette 3 and wells 3A, 3B and 3C are PCR reactions amplifying the last tRNA gene of the Chr11 tRNA array. A summary of the primers used and amplicon sizes may be found in **Table 4.5**. (**D**) Colony PCR of yeast isolates to screen for the SC.t10.07 tRNA cassette on the Chr10 tRNA array (illustrated in **Figure 4.14**; n = 22). The expected amplicon size is 489 bp. Wells A4, B6, B7 and B8 produce a band corresponding to the SC.t10.07 tRNA cassette. Well B11 is a positive control and well B12 is a negative control. YCp2937 and YCp2938 primers were used and may be found in **Appendix VI**. (**E**) Restriction digest verification of plasmids extracted from yeast isolates, transformed into *E. coli* and subjected to miniprep (n = 4). Plasmid DNA was digested using the SpeI restriction enzyme to produce expected fragment sizes of 13,917 bp, 7,319 bp, 1,896 bp, 557 bp and 61 bp. The coordinates for each restriction enzyme cut site are located on the plasmid map in **Figure 4.14**. The predicted gel pattern and fragment sizes of the 2-log ladder is shown on the right. (**F**) Sanger sequencing verification of the Chr10 tRNA array in the *LEU2* inchworm vector using the M13R (YCp3002) and YCp3100 primers (n = 1). Sequencing reads are indicated by red arrows.

4.3.4. Inchworming the Chr15 tRNA Array into the Neochromosome

The integration of the Chr15 tRNA array (the final round of inchworming) is shown in **Figure 4.18** as a representative example of the inchworming process used to construct the neochromosome. This process is illustrated in **Figure 4.16**. It should be noted that

other rounds of inchworming included variations of the following approach. The Chr15 tRNA array inchworm fragment contained the *URA3* selective marker and the necessary homology regions to replace the existing *LEU2* marker adjacent to the Chr14 tRNA array on the growing neochromosome.

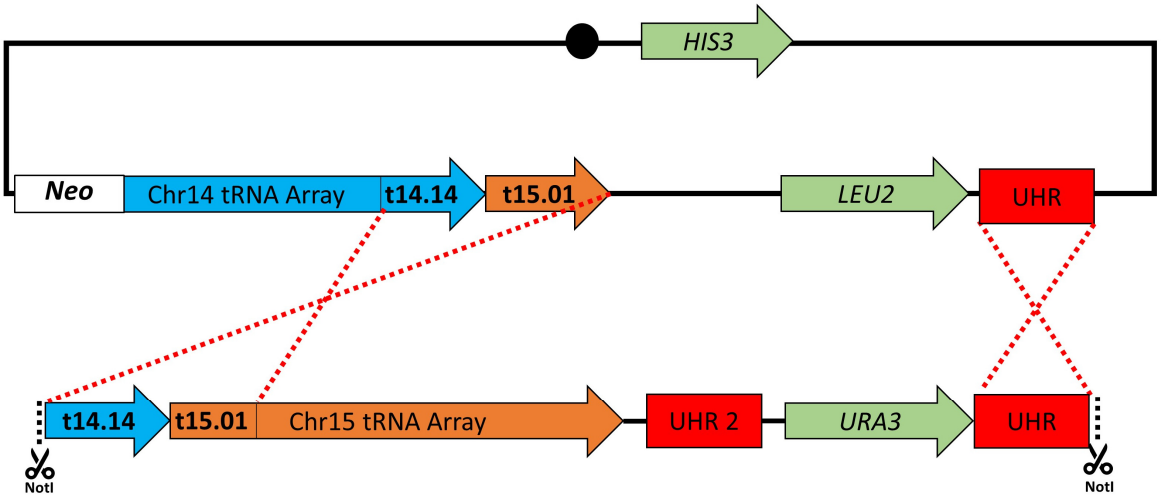


Figure 4.16: Schematic representation describing the final round of inchworming. The figure illustrates the process of introducing the 12,903 bp Chr15 tRNA array into the 176,162 bp tRNA neochromosome adjacent to the Chr14 tRNA array, mediated by two homology arms and the replacement of the *LEU2* marker with the *URA3* marker. The black dot indicates the centromere, tRNA arrays are indicated by coloured arrows, auxotrophic markers by green arrows and the 500 bp UHR with a red box. A second UHR (UHR 2) is integrated to facilitate subsequent removal of the *URA3* marker. In the above image, the white *Neo* box indicates the main body of the tRNA neochromosome sequence (not shown to scale). The NotI restriction sites used to recover the Chr15 inchworm sequence (15,840 bp) are indicated by scissors and black dotted lines. In this instance, the blue 656 bp 't14.14' is the last tRNA cassette of the Chr14 tRNA array and the orange 644 bp 't15.01' is the first tRNA array of the Chr15 tRNA array. These homologous elements, together with the 500 bp UHR sequence, were designed to recombine and facilitate integration of the Chr15 tRNA array.

The Chr15 inchworm fragment was first recovered from the *URA3* inchworm vector using NotI (**Figure 4.18A**) prior to gel purification and transformation into yeast housing the growing tRNA neochromosome (**Figure 4.18B**). A colony PCR was then performed using

primers designed to amplify across the homologous recombination junction region. Colony PCR revealed that 39 out of 94 (or ~41%) of isolates potentially tested positive for successful integration of the Chr15 tRNA array (**Figure 4.18C**). The location and design of these colony PCR primers is illustrated in **Figure 4.17A**.

During neochromosome construction, isolates were regularly observed to house all three *LEU2*, *URA3* and *HIS3* selective markers (see **Figure 4.21** as a representative example), suggesting that the neochromosome existed in the cell with a copy number greater than one. To induce loss of the unwanted secondary marker (in this case *LEU2*), it was necessary to re-streak isolates for single colonies followed by replica plating onto SC-His, SC-Leu and SC-Ura to identify isolates with the correct phenotype. **Figure 4.18D** displays a representative example of six isolates subjected to induced marker loss (although it should be noted in this instance that most cells had already lost their *LEU2* marker prior to re-streaking). A total of 36 isolates were re-streaked and replica-plated in this manner.

Single isolates displaying the correct phenotype were then screened for the presence of the SC.t07.17 and SC.t12.16 tRNA cassettes (**Figure 4.18E**) on the Chr7 and Chr12 tRNA arrays, respectively, to quickly verify the presence or absence of the neochromosome. The location of these primers is illustrated on **Figure 4.17B**. Out of 35 isolates tested, 5 tested positive for the presence of the Chr7 and Chr12 tRNA arrays, indicating a significant loss of neochromosome DNA in the majority of isolates tested. To further verify neochromosome structural integrity, these isolates were then screened using colony PCR for the presence of six tRNA cassettes (on the Chr6, Chr2, Chr16, Chr10, Chr13 and Chr14 tRNA arrays) scattered throughout the neochromosome. The location of these primer binding sites is illustrated in **Figure 4.17B** and **Table 4.6**. Four out of five isolates tested positive for all six tRNA arrays (**Figure 4.18F**). Full PCR tag analysis was then performed on one isolate to verify successful integration and ensure the presence of all tRNA cassettes on the neochromosome (**Figure 4.19**).

The results show successful integration of the Chr15 tRNA array following the final round of inchworming. Out of 94 isolates screened in **Figure 4.18**, four eventually tested positive for successful integration in addition to the presence of the neochromosome, indicating an overall efficiency of approximately 4%. During the latter rounds of inchworming, the efficiency of successful integration events decreased significantly and necessitated numerous steps to identify correct integration. However, the process overview described in this section was shown to successfully identify an isolate housing the complete circular neochromosome.

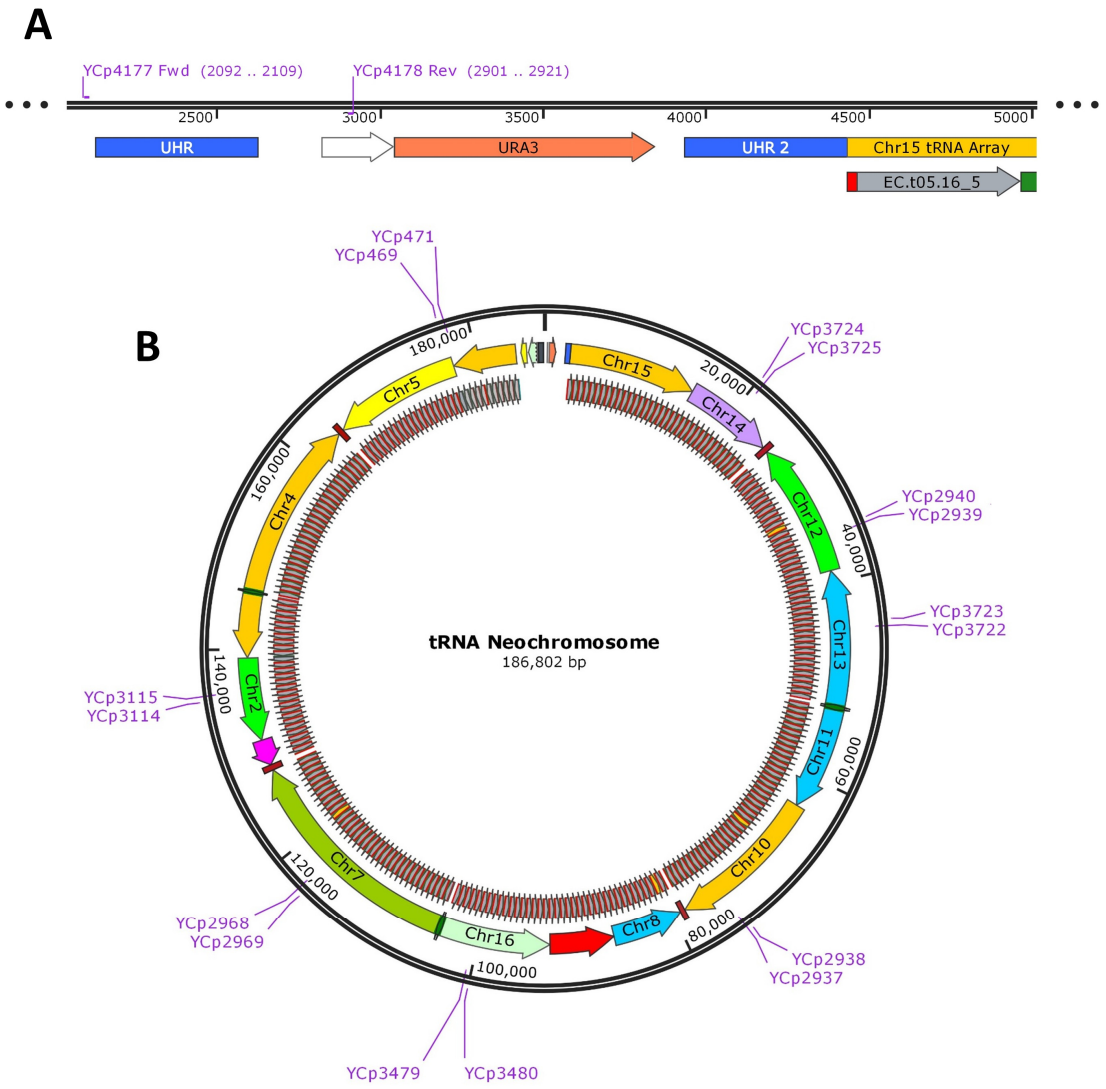


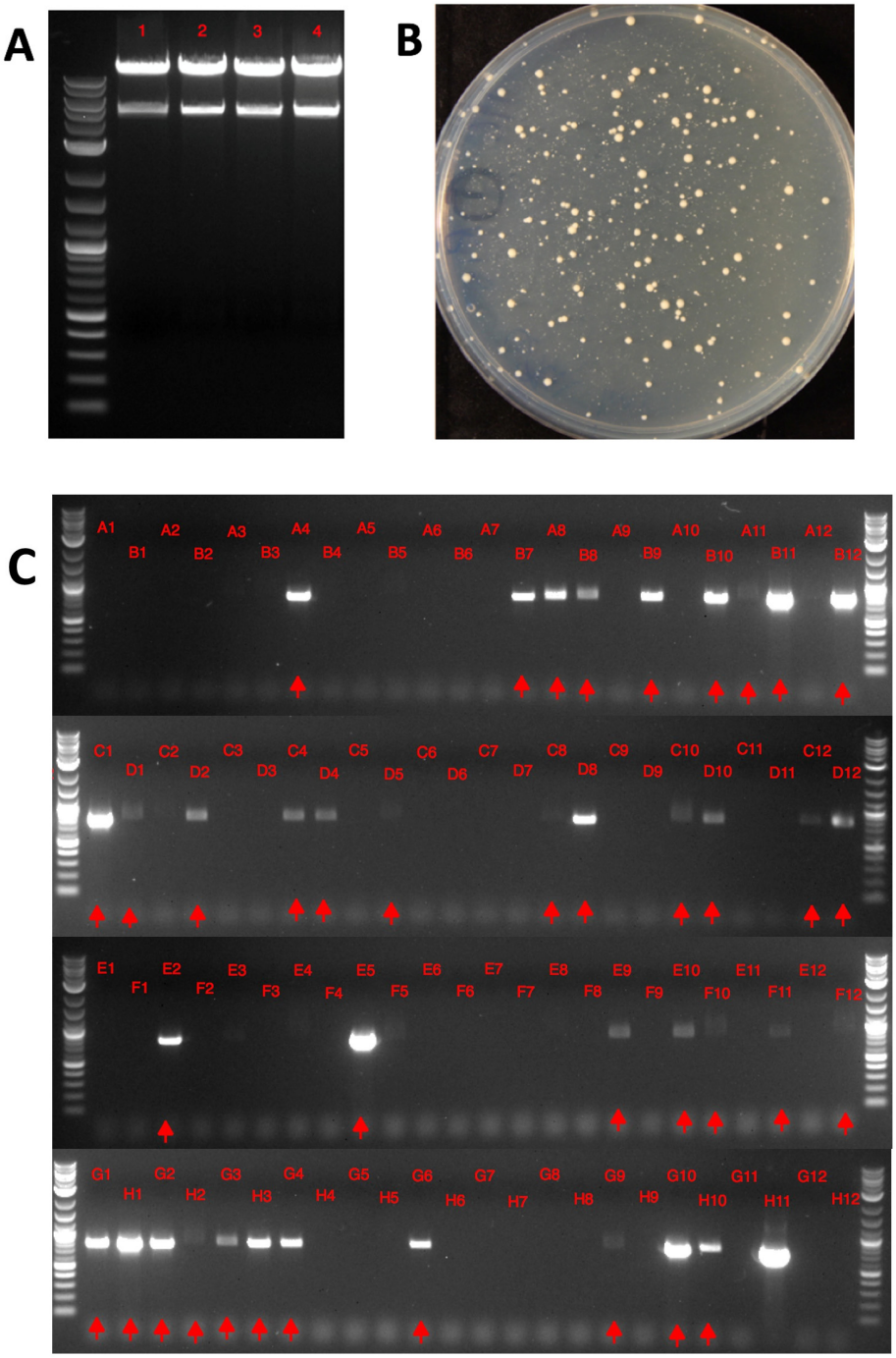
Figure 4.17: Graphical overview of the location of primers used to verify the final round of inchworming. The above images are intended to illustrate the location of primer binding sites on the body of the tRNA neochromosome structure. The location of **(A)** in relation to the rest of the neochromosome is at the 12 O'clock position on **(B)**. In all cases above, primers are denoted by pink lines (the sequences may be found in **Appendix VI**). **(A)** Primers used to amplify across the inchworming junction region. YCp4177 was designed to anneal to the left junction on the tRNA neochromosome sequence and YCp4178 was designed to anneal internally to the Chr15 inchworm fragment. The expected amplicon size is 830 bp (see **Figure 4.18C**). The three dots denote the remainder of the tRNA neochromosome sequence.

[Figure legend continued on following page]

Figure 4.17 (Continued) (B) Overall map of the circular tRNA neochromosome noting the location of primer binding sites used to verify the presence of the tRNA neochromosome structure. A full list of the primers used may be found in **Table 4.6**, primer sequences may be found in **Appendix VI**.

Table 4.6: Summary of primers used to verify the presence of the tRNA neochromosome structure. Each primer was designed to anneal with the 5' and 3' flanking sequences of each tRNA cassette in order to verify the presence of each tRNA array. A full list of the primer sequences may be found in **Appendix VI**.

tRNA Array	tRNA Cassette	Forward Primer	Reverse Primer	Expected Amplicon Size (bp)
Chr7	SC.t07.17	YCp2968	YCp2969	499
Chr12	SC.t12.16	YCp2939	YCp2940	485
Chr6	SC.t06.01	YCp469	YCp471	515
Chr2	SC.t02.08	YCp3114	YCp3115	480
Chr16	SC.t16.12	YCp3479	YCp3480	447
Chr10	SC.t10.07	YCp2937	YCp2938	489
Chr13	SC.t13.09	YCp3722	YCp3723	554
Chr14	SC.t14.07	YCp3724	YCp3725	483



[Figure continued on following page]

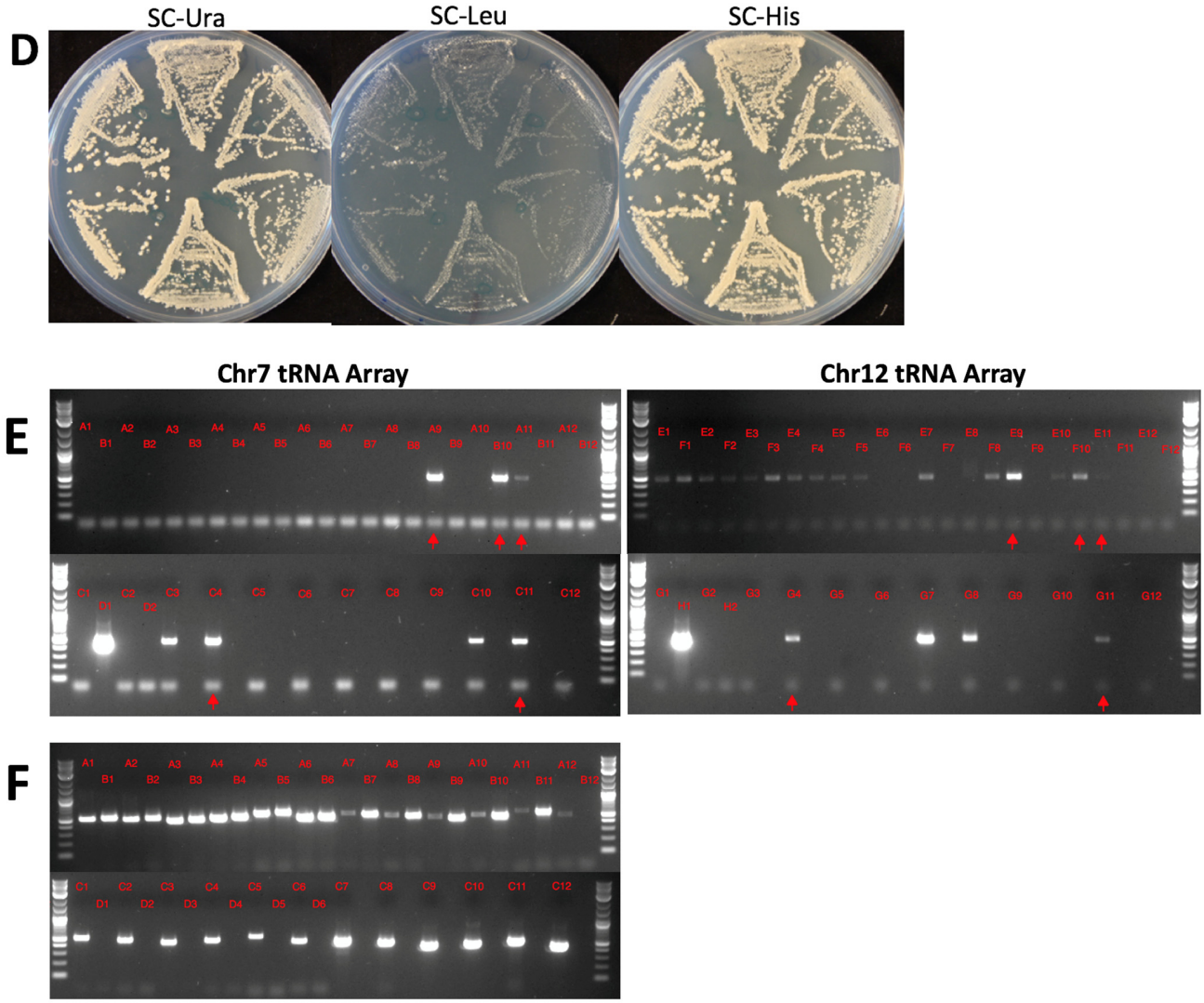


Figure 4.18: Representative results describing the final round of neochromosome construction in BY4741. The figure describes the individual steps used to introduce the 15,840 bp Chr15 tRNA array into the BY4741 strain background housing the growing neochromosome using the inchworming method. An illustration of this process may be found in **Figure 4.16**. A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1**. (**A**) Recovery of the Chr15 inchworm sequence containing the *URA3* marker from its growth vector (n = 4). Restriction digests with NotI were undertaken in replicates and later pooled together to maximise quantity of DNA recovered from the gel. The expected restriction digest fragment sizes are 15,840 bp and 4,994 bp. The upper bands were gel excised and purified.

[Figure legend continued on following page]

Figure 4.18 (Continued) (**B**) Representative yeast colonies on SC-Ura following transformation of the inchworm fragment into BY4741 housing the previous Chr14 tRNA array. (**C**) Colony PCR with primers designed to screen the homologous recombination junction region (illustrated in **Figure 4.17A**) using YCp4177 and YCp4178 to produce an expected amplicon size of 830 bp (n = 94). Each well represents a single colony tested. Well H11 is the positive control and well H12 is the negative control. Red arrows indicate isolates potentially testing positive for integration. (**D**) Re-streaked and replica plated isolates following inchworming (n = 36). In this instance, most isolates have completely lost the unwanted *LEU2* marker prior to re-streaking. (**E**) Colony PCR to screen for the presence of the SC.t07.17 and SC.t12.16 tRNA cassettes located on the Chr7 and Chr12 tRNA arrays (n = 36). The primers used may be found on **Table 4.6**, and produce expected amplicon sizes of 499 bp and 485 bp. The two gels in this case are staggered i.e. yeast isolate 1 was screened for the Chr7 tRNA array in well A1 and the Chr12 tRNA array in well E1. Well D1 and D2 are the positive and negative controls for the Chr7 tRNA array, respectively, and well H1 and well H2 are the positive and negative controls for the Chr12 tRNA array, respectively. Red arrows indicate candidate isolates that tested positive for both tRNA arrays. (**F**) Colony PCR to screen for the presence of six tRNA cassettes (located on the Chr6, Chr2, Chr16, Chr10, Chr13 and Chr14 tRNA arrays; n = 5). PCR reactions are grouped into sets of six, i.e. wells A1 to A6 is isolate 1, wells A7 to A12 is isolate 2 *etc*. A full list of the tRNA cassettes screened, primers used and amplicon sizes may be found on **Table 4.6**. Wells C7 to C12 are positive controls, wells D1 to D6 are negative controls. Isolate 4 (wells B7 to B12) is missing a band on well B12 (the Chr14 tRNA array).

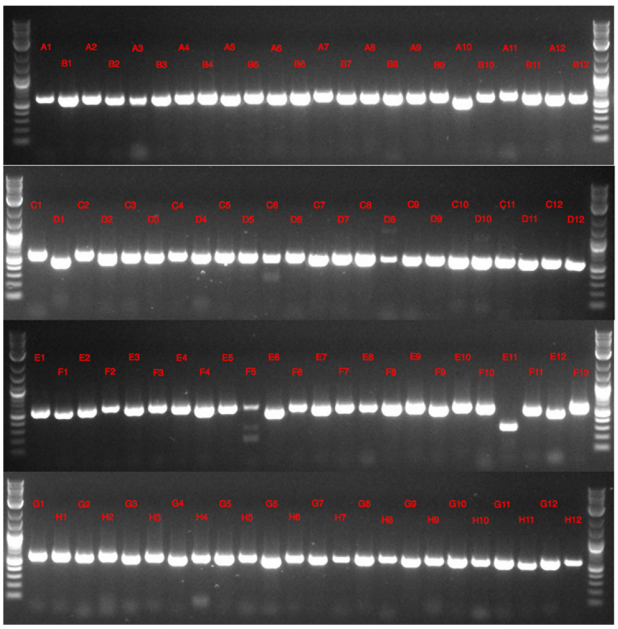
4.3.5. Neochromosome Verification: Full PCR Tag Analysis and Negative Control

Primers were designed to anneal with the 5' and 3' flanking sequence of each tRNA cassette, with the presence of a corresponding band on an agarose gel confirming the presence of each tRNA gene. The neochromosome was constructed independently in two strain backgrounds: **Figure 4.19** displays the results of a representative PCR tag assay following the final round of inchworming in BY4741. **Figure 4.20** displays a negative control of all PCR tags performed on a BY4741 strain housing an empty pRS413 plasmid. The full list of PCR tags corresponding to each well may be found in **Table II** in Appendix III.

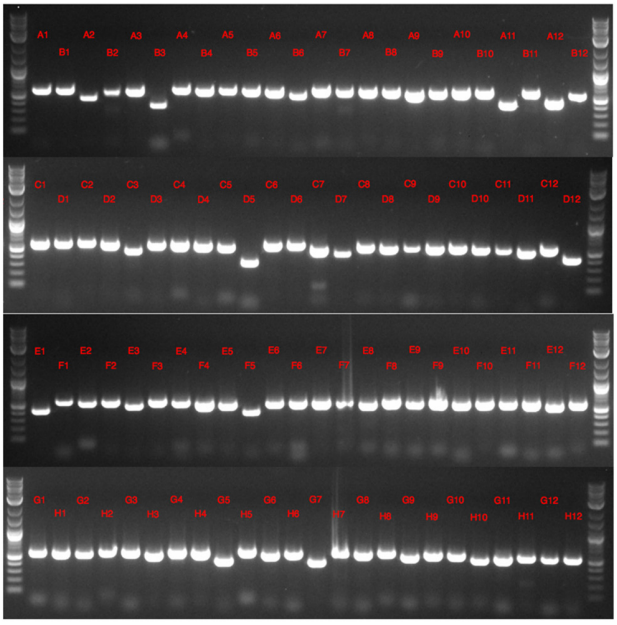
The results show that PCR tags are a reliable method of detecting the presence of individual tRNA cassettes. The non-specific primer annealing and amplification observed

for the PCR tag negative control (**Figure 4.20**) did not adversely affect the screening process used to verify neochromosome constructs.

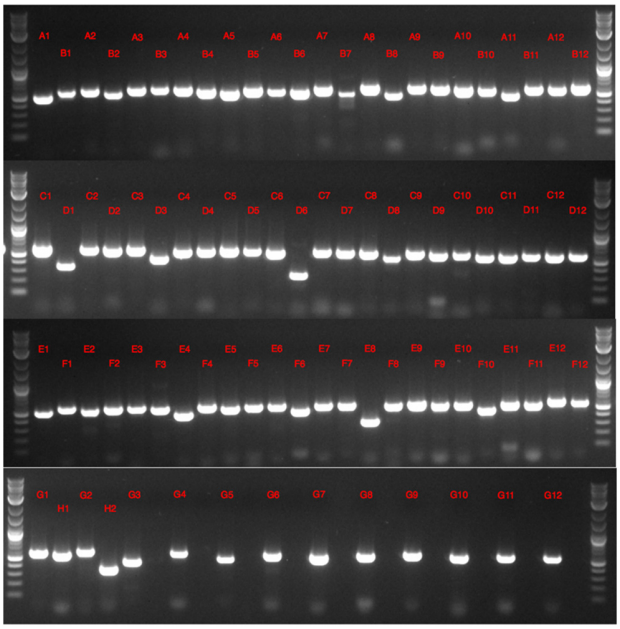
Gel 1:



Gel 2:



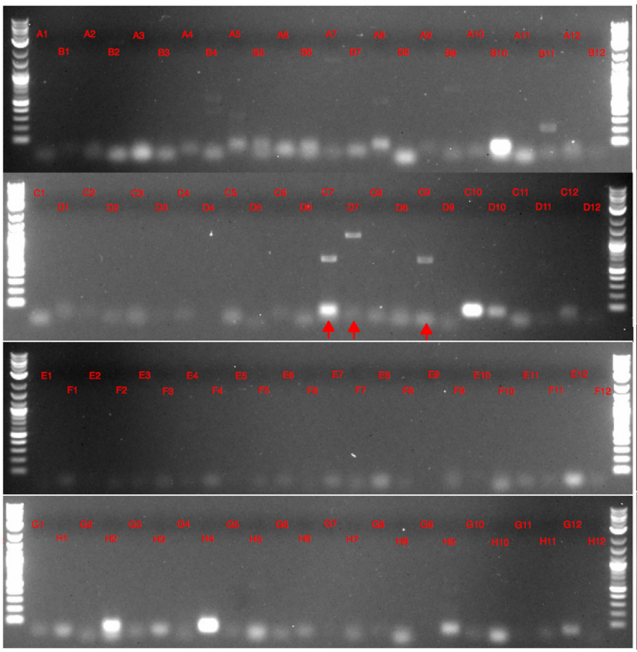
Gel 3:



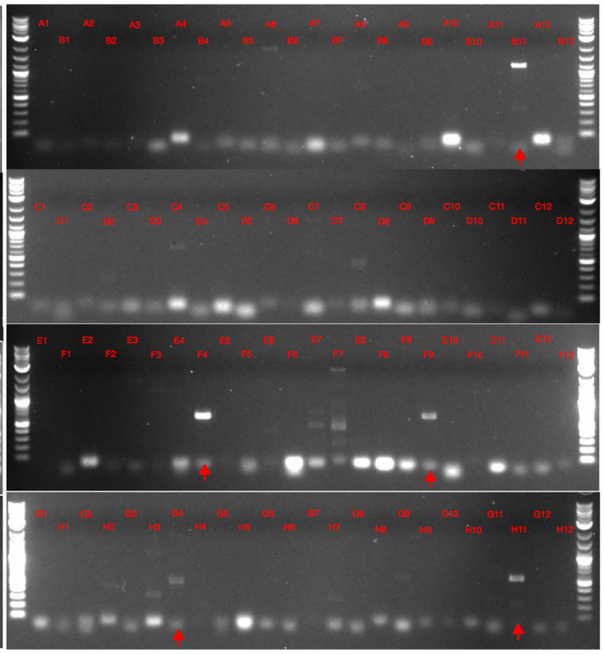
[Figure legend on following page]

Figure 4.19: Full PCR tag analysis of the intact neochromosome following the final round of inchworming in BY4741. The figure illustrates the verification of the presence of each individual tRNA cassette of the tRNA neochromosome on one isolate. Each well on the above three gels correspond to PCR reactions performed for individual tRNA cassettes – the full list of primers used may be found in Appendix III, **Table II**. A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1**. Gel 1, Gel 2 and Gel 3 correspond to PCR tag Plate 1, Plate 2 and Plate 3, respectively. Some primer pairs initially failed to produce a PCR product, and were later re-designed to produce smaller amplicon sizes – these may be observed as smaller band sizes on the above gel.

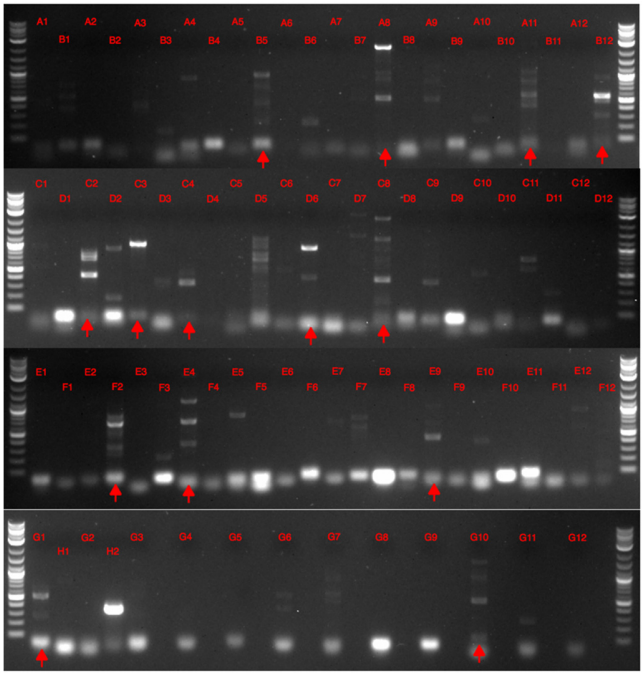
Gel 1:



Gel 2:



Gel 3:



[Figure legend on following page]

Figure 4.20: Negative control of the PCR tags used to verify neochromosome construction. The above PCR reactions were performed using the PCR tag primer pairs on a BY4741 strain housing an empty plasmid (pRS413; n = 1). Red arrows indicate primer pairs that produce more significant non-specific amplification. Well H2 is a positive control designed to amplify *AmpR* on pRS413. Each well on the above three gels correspond to PCR reactions performed for individual tRNA cassettes – the full list of primers used may be found in Appendix III, **Table II**. A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1**. Gel 1, Gel 2 and Gel 3 correspond to PCR tag Plate 1, Plate 2 and Plate 3, respectively.

4.4. Discussion

This chapter presented a step-by-step overview of the process used to construct the tRNA neochromosome and described the following approaches:

1. Combining split tRNA arrays from their constituent parts;
2. Construction of a series of vectors used to facilitate inchworming;
3. The introduction of tRNA arrays into these vectors;
4. A process overview of neochromosome construction and verification itself.

4.4.1. General Approach to tRNA Neochromosome Construction

Splitting tRNA arrays into sequences less than 10 kb was intended to reduce the overall time required for DNA synthesis from external vendors. These tRNA array parts were required to be later combined, which was undertaken by PCR amplifying the pRS413 backbone with primers designed to contain homology for subsequent Gibson Assembly. However, it was observed that Gibson Assembly failed to produce the desired construct for some tRNA array constructs. A solution was obtained by using *in vivo* homologous recombination in BY4741 to combine each problematic tRNA array. This approach was observed to be successful on every attempt.

The construction of a series of universal inchworm vectors provided a means to build both current and future versions of the neochromosome. These inchworm vectors provide the flexibility to introduce any sequence of DNA required into a growing neochromosome (notably the >23 kb Chr7 tRNA array). It was later recognised that the redundant presence of a *HIS3* marker on each inchworm vector introduces the risk of generating false-positives, and so may be omitted in future constructs.

Traditional ligation or more modern strategies like Golden Gate (Engler *et al.*, 2008) were avoided when introducing tRNA arrays into inchworm vectors (Section 4.2.4). This is because of the tendency of commercial restriction enzymes to cut internally into tRNA array sequences: of the few restriction enzymes available, only combinations of NotI and FseI avoided this issue. Gibson Assembly and *in vivo* homologous recombination in BY4741 were used *in lieu* of ligation, as these approaches were found to successfully remove remnant restriction sites on the ends of each tRNA array. This approach had the advantage of allowing the NotI or FseI restriction sites to be used more than once.

4.4.2. Inchworming Workflow Development

The original intention for neochromosome construction was to simply select and counter-select isolates for the correct *LEU2* or *URA3* phenotype following each round of inchworming. However, a series of challenges were observed during neochromosome construction.

After each round of inchworming, it was soon observed that isolates regularly grew on SC-His, SC-Leu and SC-Ura (**Figure 4.21**). The presence of all three auxotrophic markers (*LEU2, URA3* and *HIS3*) suggests that the neochromosome existed in the cell with a copy number greater than one. For example, if a *URA3*-based inchworm fragment was transformed into a cell containing multiple neochromosome copies housing the *LEU2*

selective marker, probability dictates that the *URA3* inchworm fragment will only integrate into one of these copies. Due to the mitotically-stable nature of the centromere, both *LEU2* and *URA3* neochromosome variants may transfer into daughter cells during mitosis.

To mitigate this phenomenon, a solution was to induce loss of this secondary selective marker by re-streaking isolates to produce single colonies. These were then subsequently replica-plated onto SC-His, SC-Leu and SC-Ura: only single colonies displaying the correct phenotype were selected for subsequent analysis (illustrated in **Figure 4.21**).

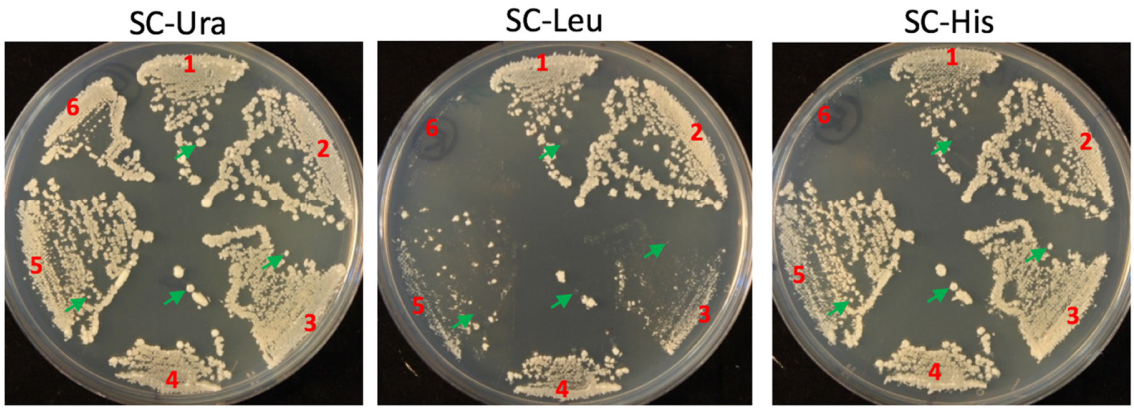


Figure 4.21: Representative example of inchwormed neochromosome isolates displaying the presence of remnant auxotrophic markers. The figure shows the process used to identify isolates displaying the correct phenotype after the inchworming process. Re-streaked isolates were replica-plated onto SC-Ura, SC-Leu and SC-His to produce single colonies for subsequent analysis. Isolate numbers subjected to re-streaking are indicated on the above plates. Green arrows indicate single colonies displaying the correct phenotype. In this instance, the *LEU2* selective marker was to be replaced by the *URA3* marker, although the remnant presence of the *LEU2* marker interfered with subsequent analysis and had to be removed by selection by induced marker loss. Only single colonies displaying the correct phenotype were subjected to further analysis. In this instance isolate 2 and isolate 6 were omitted from subsequent screening.

The second challenge observed during construction was a loss of neochromosome DNA. A significant proportion of isolates housing large deletions were regularly observed after each round of inchworming, with the number of isolates housing a deletion increasing proportionally to neochromosome size (illustrated in Section 4.3.4). For the latter stages of neochromosome construction, a workflow was developed to ensure that each isolate displays the expected phenotype, reliably verify successful integration of each tRNA array and ensure the presence of the full neochromosome sequence. This process workflow is summarised as follows:

1. Colony PCR using primers designed to amplify across the junction of each homologous region were performed to verify integration of each tRNA array.
2. Isolates were re-streaked onto selective media to induce loss of the unwanted secondary selective marker and to produce single colonies. These plates were then replica-plated onto SC-His, SC-Ura and SC-Leu (see **Figure 4.21**).
3. Single colonies displaying the correct phenotype were subjected to an initial round of colony PCR screening using primers designed to amplify one or more tRNA cassettes.
4. A further round of screening using primer pairs designed to amplify tRNA cassettes on the Chr6, Chr2, Chr16, Chr10, Chr13 and Chr14 tRNA arrays (or variations thereof) was performed to verify the presence of an intact neochromosome structure.
5. Full PCR tag analysis was then performed to verify the presence of all tRNA genes of the neochromosome (see **Figure 4.19**).

The inchworming method of neochromosome construction and the process workflow described above was shown to reliably and seamlessly integrate each tRNA array into the growing neochromosome. Although time consuming and inefficient during the latter rounds of construction, the screening methods described above demonstrated the feasibility of approaches used. Overall, the inchworming method may form the basis for construction of future neochromosomes.

Chapter 5: Characterising the tRNA Neochromosome

5.1. Introduction

This chapter describes the experimental work performed to characterise the tRNA neochromosome. This work is largely split into two primary themes: characterisation of individual functional components and physical manipulation of the neochromosome structure.

The former theme is of importance in determining whether genetic elements of the neochromosome functioned as expected. This work included characterisation of elements associated with neochromosome replication (ARS elements and *Ter* sites) and three, single-copy tRNA genes (*SUP61*, *TRT2* and *TRR4*). The latter included work on the physical manipulation of the neochromosome structure, including direct visualisation using pulsed-field gel electrophoresis (PFGE) and chemical extraction and transfer of the neochromosome from one cell to another. These elements of characterisation were required to verify the presence of the neochromosome as a physical, intact object. Additionally, the physical transfer of the circular neochromosome from one cell to another is of importance to the Sc2.0 consortium. Finally, a plate reader growth assay was performed to determine the growth rate of strains housing the neochromosome.

5.2. Characterising Neochromosome Replication Elements

A series of orthogonal origins of replication and replication termination sites were incorporated to define the profile of DNA replication associated with the tRNA neochromosome. This section describes the characterisation of these elements. It should be noted that the following results are considered qualitative and preliminary before

more in-depth characterisation methods are performed, such as that of deep-sequencing (Muller *et al*., 2014).

5.2.1. Experimental Approach

5.2.1.1. Characterising Origins of Replication

The neochromosome is designed to house four origins of replication (three from *C. glabrata* and one adjacent to the centromere). A simple plasmid replication assay was performed to test functionality of origins from *C. glabrata* by sub-cloning each into a pRS403 vector, followed by transformation into BY4741. pRS403 contains a *HIS3* auxotrophic marker but itself lacks an origin of replication and so cannot replicate in yeast without the insertion of a functional origin. Three ARS elements were recovered from their bacterial backbone through restriction digest (*chrF-444*: PspOMI + SpeI, *chrL-615*: PspOMI + SacI, *chrM-794*: EcoRI + SpeI) and ligated into pRS403 linearised with the same enzymes. Following verification through restriction digest DNA fingerprinting, 150 ng of each vector were then transformed into BY4741, serially diluted by a factor of 1 in 1,000, and plated onto SC-His media (**Figure 5.2**). The empty pRS403 vector was used as a negative control. This experiment was performed in singlicate. Plates were incubated at 30^{o}C until the appearance of sufficiently large colonies.

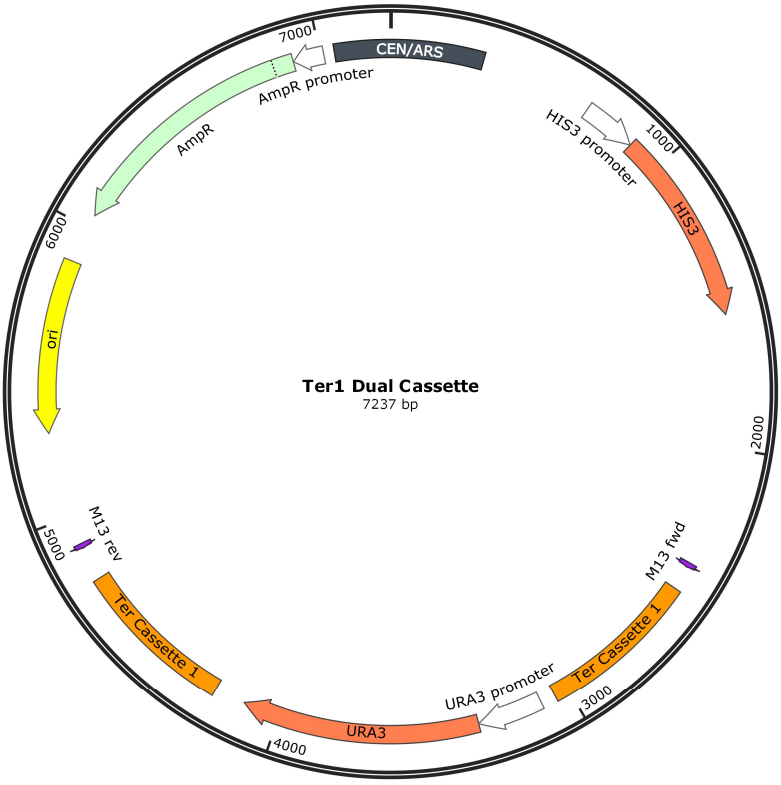
5.2.1.2. Characterising Replication Termination Sites

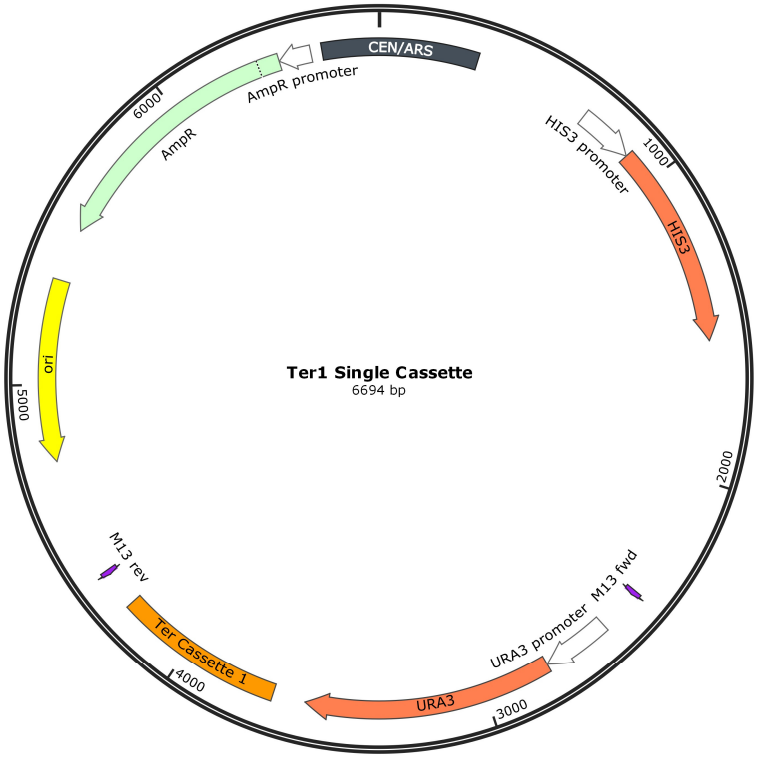
A series of four synthetic, orthogonal replication termination sites were designed by Prof. Jef Boeke (NYU Langone Medical Centre, New York, USA) and were designed to block replication in both orientations. To assay function of these termination cassettes, an experiment was designed to position two *Ter* cassettes on either side of a *URA3* marker on a pRS413 plasmid (**Figure 5.1**). If the *Ter* cassettes fail to block replication in even one

direction, the *URA3* marker would still replicate and allow growth on SC-Ura media. If *Ter* cassettes do successfully block replication in both directions, plasmids would fail to replicate the *URA3* marker and would not grow on SC-Ura media. A series of six constructs were generated, and are summarised as follows:

1. *Ter1* double termination cassette.
2. *Ter2* double termination cassette.
3. *Ter3* double termination cassette.
4. *Ter4* double termination cassette.
5. *Ter1* single termination cassette (control).
6. *URA3* inchworm vector (control).

The first four constructs were designed to test the function of each *Ter* cassette. The *Ter1* single termination cassette was a control designed to determine whether the presence of a single termination cassette disrupts plasmid stability, and the *URA3* inchworm vector was a positive control. Each cassette was constructed by PCR-amplifying their constituent parts using primers designed to house 40 bp overhangs to a pRS413 plasmid linearised with EcoRV. Following Gibson Assembly, each cassette was verified using restriction digest DNA fingerprinting and Sanger sequencing. Each cassette was then transformed into BY4741, serially diluted onto SC-His agar to produce single colonies, and replica-plated onto both SC-His and SC-Ura agar (**Figure 5.3**). Colonies that fail to grow on SC-Ura are presumed to have lost the *URA3* marker through successful blocking of replication.





[Figure legend on following page]

Figure 5.1: Representative plasmid maps describing the general design of constructs used for the replication termination assay. The above plasmid maps are intended to illustrate the general design used for the other constructs used in the replication termination assay. The *Ter1* double termination cassette (7,237 bp) is shown on the top and the *Ter1* single termination cassette control (6,694 bp) is shown on the bottom. The replication termination cassettes (orange boxes) are located between the two M13 primer binding regions. The intervening *URA3* marker is indicated by an orange arrow. Standard functional elements of the pRS413 shuttle vectors are indicated on the above diagrams. The other constructs were designed as variants of the *Ter1* double termination cassette.

5.2.2. Results

5.2.2.1. Replication Origins from *C. glabrata* Support Plasmid Replication in *S. cerevisiae*

Origins of replication from *C. glabrata* were cloned into the pRS403 yeast vector, transformed into BY4741 with normalised DNA concentrations, and serially diluted by a factor of 1 in 1,000. Constructs were transformed into BY4741 in singlicate. The results show that all origins of replication support plasmid replication in *S. cerevisiae*, and no colonies were observed for the pRS403 negative control (**Figure 5.2**). It should be noted that ARS *chrL-615* displayed a smaller colony size, suggesting that this origin is weaker than the others.

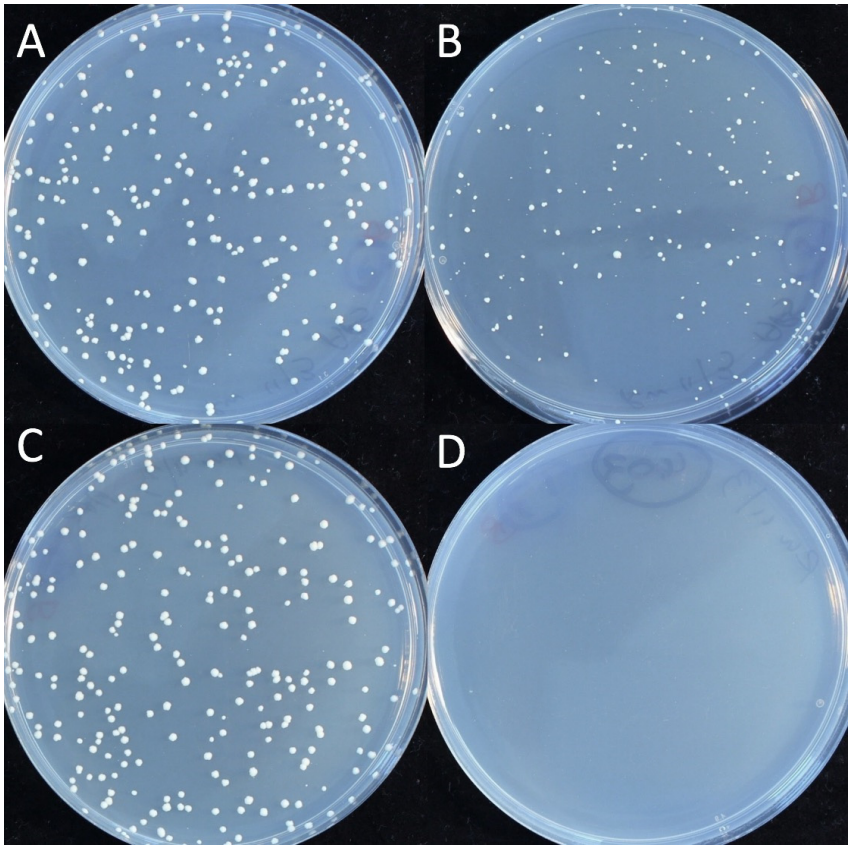
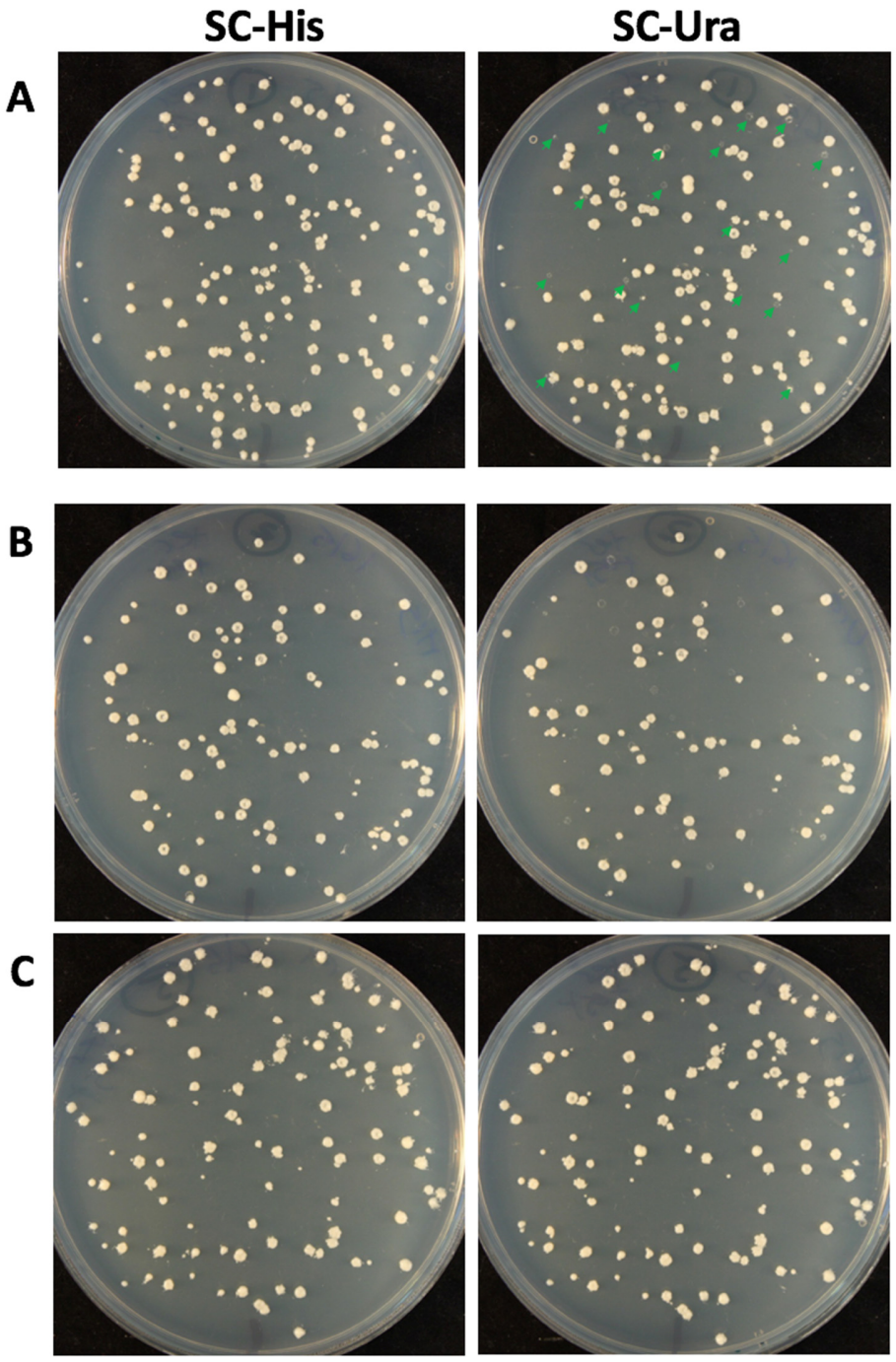


Figure 5.2: Origins of replication from *C. glabrata* support pRS403 plasmid replication in *S. cerevisiae*. The figure demonstrates the functionality of each origin of replication recovered from *C. glabrata*. Each plate houses BY4741 colonies transformed with pRS403 containing the following origins of replication: (**A**) *chrF-444*, (**B**) *chrL-615*, (**C**) *chrM-794* and (**D**) pRS403 (empty plasmid control). Strains were transformed in singlicate (n = 1).

5.2.2.2. Convergent Replication Termination Sites Block Replication of a *URA3* Marker

A series of cassettes were constructed to test function of the replication termination sites used for the neochromosome. **Figure 5.3** displays representative results for the *Ter1* and *Ter2* double replication termination cassettes, and **Table 5.1** shows the full results for all *Ter* cassettes. With the exception of the *Ter1* cassette, this experiment was performed in experimental singlicate. The results show that all four double *Ter* cassettes successfully blocked replication of the *URA3* marker for some transformants, with indicative

efficiencies of between 15% and 28%. All control cassettes were observed to retain their *URA3* marker as expected.



[Figure legend on following page)

Figure 5.3: Yeast plates displaying representative results of the replication termination assay. The above plates are shown as a representative example for each of the other replication termination cassettes, and demonstrate the functionality of the *Ter1* termination cassette. Each of the above plates house isolates that were replica-plated following transformation with each replication termination cassette. Strains were transformed as singlicate (n = 1). The *Ter1* (**A**) and *Ter2* (**B**) double termination cassettes are shown as representative examples for the other termination cassettes. The *URA3* inchworm vector (**C**) is shown as a positive control. In all cases, plates on the left are SC-His and plates on the right are SC-Ura. Colonies that fail to grow on SC-Ura may be observed as halos (denoted by green arrows for the Ter1 double termination cassette).

Table 5.1: Indicative replication termination efficiencies of each *Ter* construct. Letters A, B and C in the table below correspond to those in **Figure 5.7**. The following results were obtained from experimental singlicates (n = 1) (with the exception of the *Ter1* double termination cassette, which produced an efficiency of 19% from 58 colonies out of 306 failing to grow on SC-Ura from an experimental repeat).

	Construct	Colonies on SC-His	Colonies on SC-Ura	Replication Termination Efficiency (%)
A	*Ter1* double termination cassette	153	133	15.0
B	*Ter2* double termination cassette	96	69	28.1
	Ter3 double termination cassette	133	107	19.6
	Ter4 double termination cassette	113	90	20.4
	Ter1 single termination cassette (control)	272	272	0.0
C	*URA3* inchworm vector (control)	110	110	N/A

5.2.3. Discussion

The tRNA neochromosome incorporates orthogonal elements for a defined profile of DNA replication. Prior to more in-depth characterisation methods are performed in future, it was necessary to characterise the orthogonal origins of replication and replication *Ter* sites. It should be noted that, although both experiments were intended

to be qualitative and not quantitative, they were performed as singlicates (*i.e.* only one construct was tested at a time). For further confidence in the above results, they may be performed with replicates in future.

Characterisation of origins of replication from *C. glabrata* demonstrated that they all functioned in *S. cerevisiae* (**Figure 5.2**), although one origin (ARS *chrL-615*) displayed a smaller colony size, indicating that this origin is weaker than the others. The assay used to characterise *Ter* sites shows that they all successfully blocked replication of a *URA3* marker (**Figure 5.3** and **Table 5.1**), although the results indicate that the overall efficiency of fork blocking may not high enough to completely disrupt replication of the *URA3* marker.

5.3. **Individual-Level tRNA Characterisation**

The synthetic yeast genome will eventually have all 275 tRNA genes removed and relocated onto a tRNA neochromosome. However, it is unclear what systematic effects this may have on cellular processes due to the presence of non-*S. cerevisiae* flanking sequences and the extra-genomic context of tRNA expression. To assay tRNA functionality on an individual level, three, single-copy tRNA genes were investigated: *SUP61* (*tS(CGA)C)*, *TRT2* (*tT(CGU)K*) and *TRR4* (*tR(CCG)L*). All three tRNA genes are essential in *S. cerevisiae*. These tRNA genes lie on ChrIII, ChrXI and ChrXII, respectively, and were selected as these three chromosomes were either fully-synthesised (SynIII), or close to completion (SynXI and SynXII) at the time of performing the experiment.

5.3.1. *Experimental Approach*

As this characterisation work was undertaken prior to synthesis of neochromosome DNA, individual synthetic *SUP61, TRT2* and *TRR4* tRNA cassettes were constructed *de novo*

using the Build-A-Genome protocol and verified with Sanger sequencing. Following ligation into a specialised vector (pPC047; illustrated in **Figure 5.4**) containing a *URA3* marker and homology arms specific to the *HO* locus of BY4741, each construct was excised from its backbone and integrated into the *HO* locus of BY4741 (Δ*sup61 +* [pRS415-*SUP61*]*,* Δ*trt2* [pRS415-*TRT2*] *and* Δ*trr4* [pRS415-*TRR4*]*)*. Genome integration was performed to prevent plasmid-copy number effects on tRNA expression. A summary of the primary steps used for this experiment may be found in **Figure 5.5**.

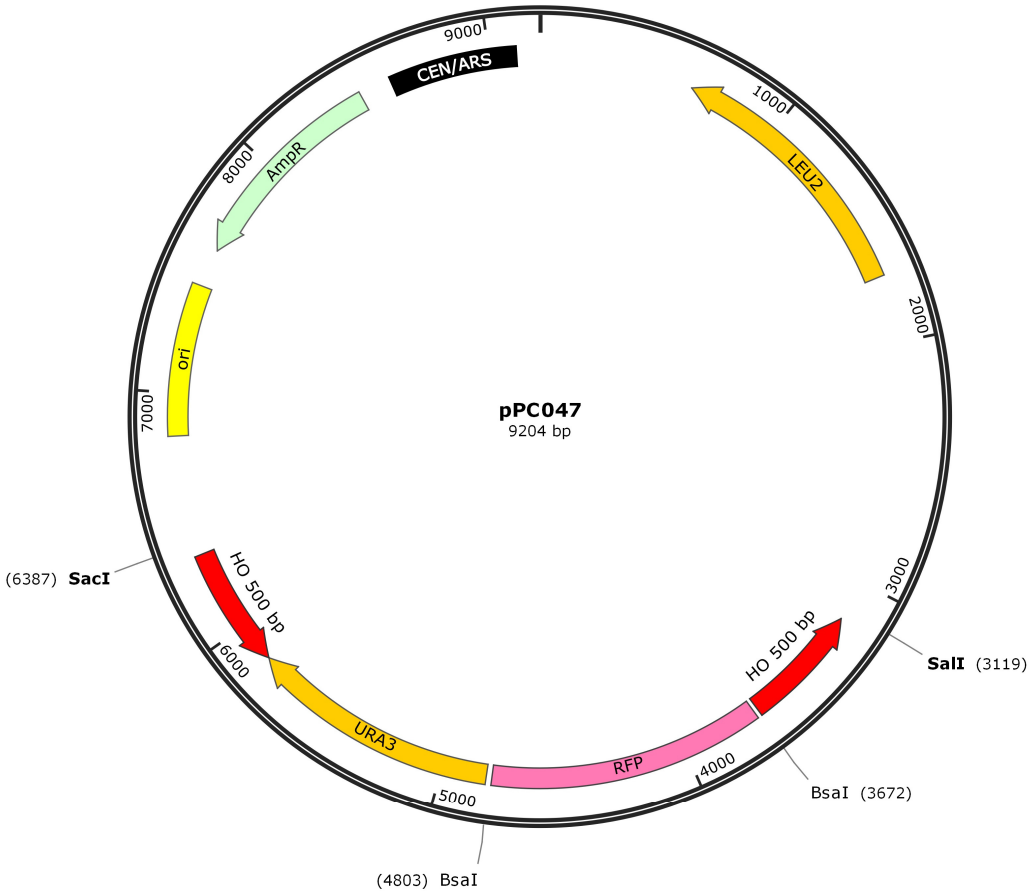


Figure 5.4: Plasmid map of pPC047. The plasmid above contains a two, 500 bp homology arms for integration into the *HO* locus of BY4741. The *URA3* marker provides a selective marker for integration into the *HO* locus. Two BsaI sites (indicated on the above plasmid map) flank a Red Fluorescent Protein (RFP) gene, providing a means to ligate a DNA sequence into the above construct. The presence of SacI and SalI restriction sites on either side of the cassette facilitates recovery of the *SUP61*, *TRT2* and *TRR4* tRNA cassettes from the above construct prior to gel extraction and transformation into the *HO* locus.

Before integration of each tRNA cassette, wild-type copies of each essential tRNA gene were removed by PCR-amplifying a *KanMX* marker with 40 bp homology arms specific to the surrounding region of each tRNA gene. The viability of each Δ*sup61,* Δ*trt2 and* Δ*trr4* strain was supported by the presence of a synthetic copy of each on a pRS415 plasmid. Following integration of *SUP61*, *TRT2* and *TRR4* into the *HO* locus, induced plasmid loss was used to remove the plasmid-based copy of each. Removal of each wild-type tRNA gene and integration into the *HO* locus (data not shown) was verified using colony PCR with primer pairs designed to amplify the junction regions of each locus. Colony PCR (**Figure 5.6B**) demonstrates the absence of the wild-type copy of each tRNA gene. Basic phenotypic characterisation was then undertaken on these strains (**Figure 5.6A**), as well as Northern blot (**Figure 5.7**). Biological duplicates were used for each experiment.

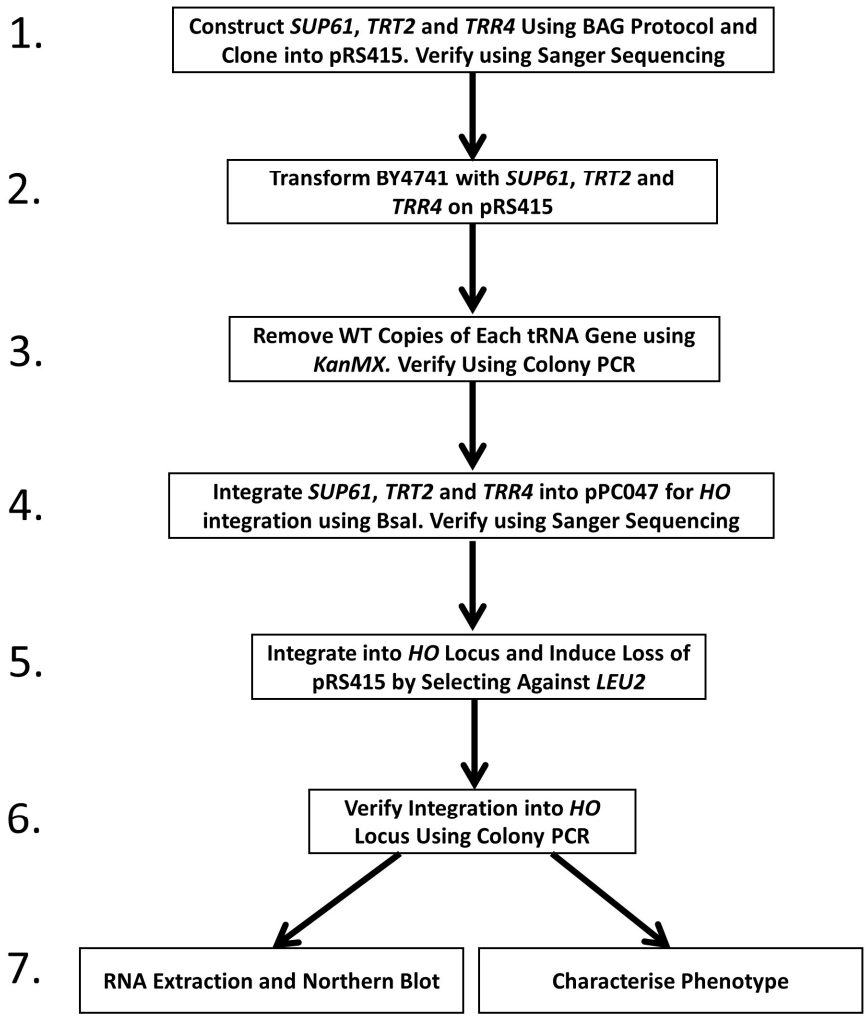


Figure 5.5: Flow diagram describing the primary steps involved for the single-copy tRNA complementation experiment. The above diagram describes the overall process of integrating synthetic copies of the *SUP61*, *TRT*2 and *TRR4* tRNA genes and the removal of wild-type copies from the BY4741 genome.

Northern blot is a highly sensitive and quantitative method to characterise tRNA expression. However, mature tRNA molecules are highly abundant (Wei, 2013), and so subtle differences between their levels are difficult to quantify. Normally, probes are designed to anneal with either the 5' or 3' junction region of precursor tRNA transcripts (i.e. pre-tRNAs) to provide a better indication of transcriptional level. However, flanking

sequences differ between synthetic and wild type tRNAs. As different precursor probes would be required for each, Northern blot results would not be cross-comparable. For this reason, only mature tRNA probes were used.

5.3.2. Results

5.3.2.1. Synthetic tRNA Cassettes Complement Loss of Essential tRNA Genes

A series of strains were generated housing a synthetic copy of *SUP61*, *TRT2* and *TRR4* in the *HO* locus of BY4741. These strains also had the wild-type copies of each removed (as described in Section 5.3.1.). Biological duplicates of these strains and a wild-type control (BY4741 + pRS416) were inoculated into 10 mL of SC-Ura and incubated with rotation overnight at 30^{o}C. Overnight cultures were then adjusted to a normalised OD_{600nm} of 0.03 in sterile PBS (phosphate buffered saline) solution, serially diluted, and spotted side-by-side onto SC-Ura selective media (**Figure 5.6A**). Plates were then incubated at 30^{o}C for three days.

The results show no observable difference in colony size or morphology for strains housing sole-synthetic copies of each tRNA gene when compared to the wild-type control. **Figure 5.6B** displays the colony PCR used to verify successful removal of each single-copy tRNA gene, and is shown to demonstrate the absence of each wild-type copy.

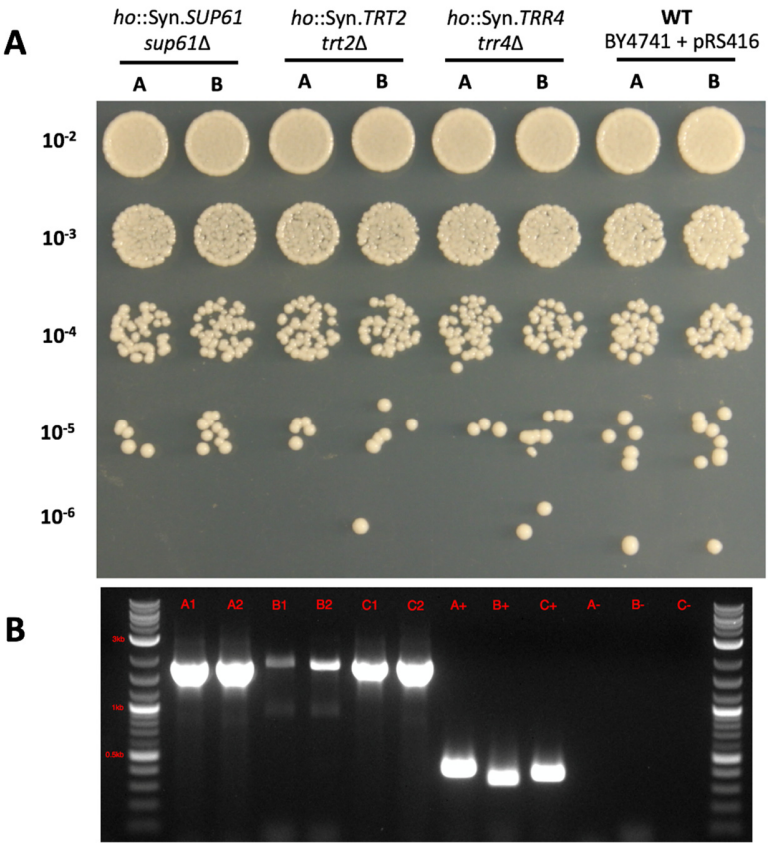


Figure 5.6: Phenotypic colony morphology of each strain housing a synthetic copy of each essential tRNA gene and colony PCR used to verify successful knock-outs. The figure demonstrates that synthetic single-copy tRNA genes can complement the absence of each wild-type copy in BY4741 with no observable effect on colony size. (**A**) Serial dilution and spotting of each strain onto SC-Ura. A and B are biological replicates (n = 2) of each strain. The dilution factor of each yeast sample is indicated on the left of the diagram. (**B**) Colony PCR performed on biological replicates (n = 2) to verify the successful removal of each wild-type copy of *SUP61* (A), *TRT2* (B) and *TRR4* (C). A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1**. Primers were designed to anneal to flanking regions around the site of integration, and so produce a 1.74 kb, 1.70 kb and 1.74 kb for successful knock-out for *SUP61*, *TRT2* and *TRR4*, respectively. Wild-type copies of each tRNA gene, however, will produce a 0.42 kb, 0.35 kb and 0.39 kb band for each respective tRNA gene. '**+** 'on the above diagram denotes wild-type control reactions (n = 1) performed on BY4741 + pRS416. '**-**' on the above diagram denotes negative control reactions (n = 1) performed using ddH_2O instead of template DNA.

5.3.2.2. Northern Blot Reveals Potential SUP61 Precursor Accumulation

No phenotypic defects were observed after spotting strains housing synthetic tRNAs side-by-side (**Figure 5.6**). To compare the mature products of synthetic and wild-type tRNA genes, Northern blot was performed using probes specific to the SUP61, TRT2 and TRR4 tRNA molecules. Biological duplicates of strains housing a sole-synthetic copy of each tRNA located in the *HO* locus were compared to strains housing wild-type copies of each. A plasmid-based copy of each synthetic tRNA was also included to investigate the effects of plasmid copy-number on tRNA expression. Each Northern blot membrane was then stripped and re-probed again for 5.8S rRNA as a normalising control.

The results (**Figure 5.7**) show that Northern blots for TRT2 and TRR4 are of poor-quality, possibly due to washes of high stringency and potential degradation of RNA, although do not appear to show significant variance in tRNA expression levels. SUP61, however, is of greater interest: an additional band can clearly be observed for synthetic copies of *SUP61*. Due to the increased size of this band relative to the wild-type tRNA, this additional band may be an accumulation of a SUP61 precursor.

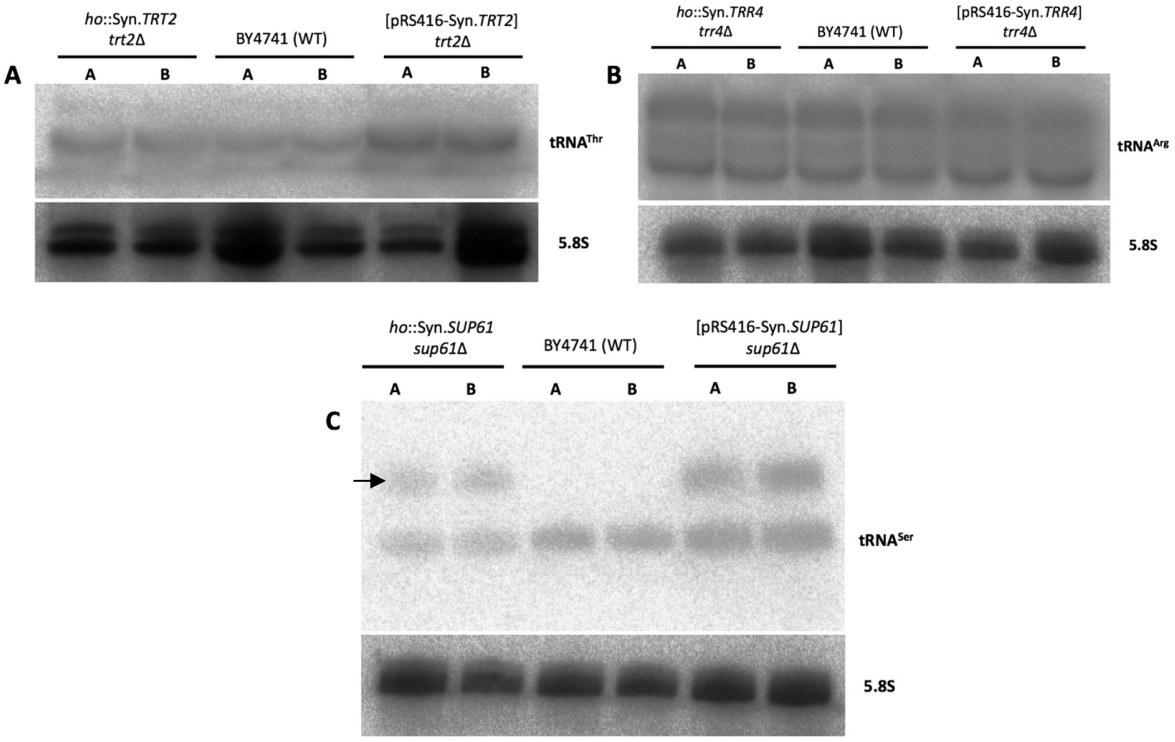


Figure 5.7: Northern blots for TRT2, TRR4 and SUP61. The figure demonstrates the presence of an additional band observed for the SUP61 tRNA. (**A**) is TRT2, (**B**) is TRR4 and (**C**) is SUP61. The first two lanes are biological replicates (n = 2) of strains housing a sole-synthetic copy of each tRNA gene in the *HO* locus. The second two lanes are wild-type controls (BY4741 + pRS416; n = 2) and the third two lanes are biological replicates (n = 2) with synthetic tRNA cassettes on a pRS416 plasmid. The black arrow indicates the presence of an additional band for the SUP61 tRNA. All samples housing a synthetic tRNA cassette have their corresponding wild-type tRNA gene removed.

5.3.3. Discussion

The neochromosome is a critical component of Sc2.0, and so it was important to ensure that synthetic tRNA genes function as expected. This could not yet be performed in a systematic manner, and so function of the essential, single-copy *SUP61, TRT2* and *TRR4* were investigated on an individual level. Wild-type copies of each gene were removed from the genome so that synthetic counterparts were required to maintain tRNA function. The results revealed no observable defect in terms of colony size (**Figure 5.6**),

demonstrating that each synthetic tRNA gene is individually sufficient to maintain viability.

However, Northern blot clearly produced an additional band for the synthetic SUP61 variant, presumably due to the accumulation of a precursor tRNA (**Figure 5.7A**). Although TRT2 and TRR4 blots were of poor quality, they do not appear to show any difference between synthetic and wild-type variants.

5.4. **Elucidating Source of Additional SUP61 Band**

An additional band can clearly be observed for the synthetic SUP61 following Northern blot (**Figure 5.7A**), presumably due to an accumulation of its precursor form. The following experiment describes the elucidation of which feature associated with the synthetic form is responsible for the additional band.

5.4.1. Experimental Approach

Following the observation of an unexpected additional band for the synthetic SUP61 tRNA, an experiment was designed to elucidate its source by generating a series of *SUP61* variants. These variants are summarised in **Figure 5.8**.

The first and second *SUP61* variants were designed to replace the synthetic 5' and 3' flanking sequences with their wild-type copies, respectively (**Figures 5.8A** and **5.8B**). The poly-T terminator signal from *A. gossypii* was observed to lie directly adjacent to the 3' end of the tRNA gene: to eliminate the possibility of a transcriptional 'readthrough' effect (Turowski *et al.*, 2016), additional bases (TTTAA) from the *S. cerevisiae SUP61* 3' flanking sequence were placed between the tRNA gene and the poly-T terminator (variant three; **Figure 5.8C**).

The synthetic copy of *SUP61* lacks an intron that is present in the wild-type copy. To eliminate the possibility of any intron-related effects on tRNA processing, the fourth variant was a wild-type copy lacking an intron (**Figure 5.8D**). The final variant was designed to investigate whether genomic location plays a role in tRNA expression or processing, and was a wild-type copy of the *SUP61* tRNA translocated to the *HO* locus (**Figure 5.8E**).

Variants 1 and 5 were synthesised *de novo* using gBlocks (IDT). The other variants were generated using PCR and nested PCR with primers designed to introduce the required changes. Each *SUP61* variant was introduced into the pPC047 vector for *HO* locus integration, verified using Sanger sequencing and integrated into the *HO* locus of BY4741 (Δ*sup61* + [pRS415-*SUP61*]). Viability of each Δ*sup61* strain was supported by the presence of a synthetic *SUP61* copy on pRS415, which was again removed by inducing plasmid loss. Integration of the *SUP61* variants into the *HO* locus was verified using colony PCR with primers designed to amplify across junction regions (data not shown). Northern blot on biological duplicates was performed as previously described (Section 5.3.1).

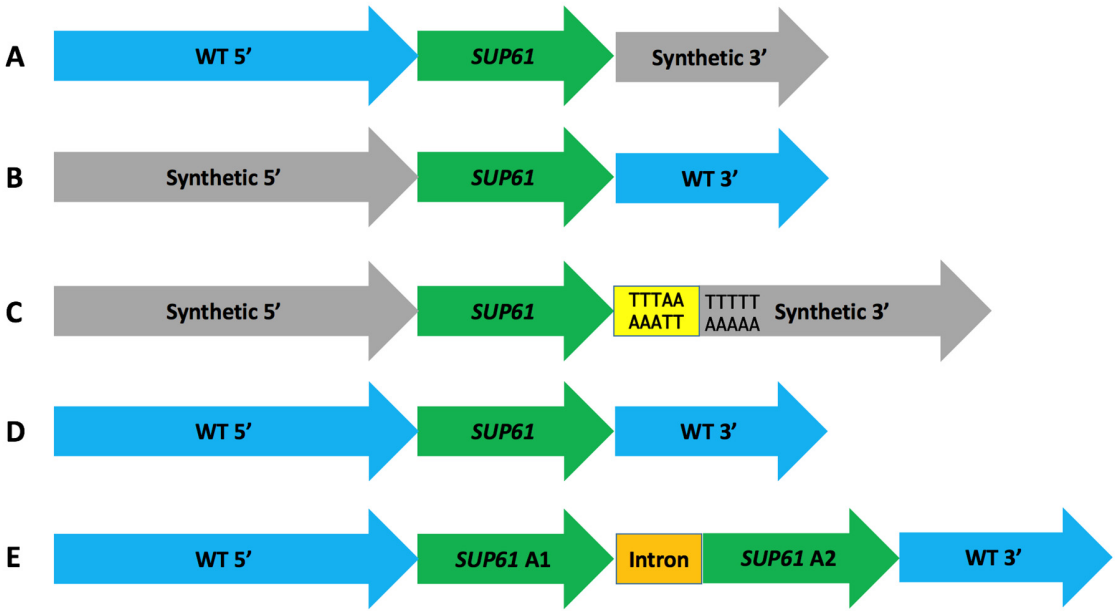


Figure 5.8: Schematic representation of *SUP61* variants designed to elucidate the source of the additional SUP61 band following Northern blot. A total of five versions of the *SUP61* gene were constructed to identify which feature was responsible for the additional band following Northern blot. tRNA genes are indicated by green arrows, wild-type flanking sequences by blue arrows and synthetic flanking sequences by grey arrows. (**A**) Synthetic *SUP61* with a wild-type 5' flanking sequence. (**B**) Synthetic *SUP61* with a wild-type 3' flanking sequence. (**C**) Synthetic *SUP61* with the presence of an additional five nucleotides (yellow box; found in the wild-type 3' flanking sequence) prior to the 3' poly-thymidine terminator. (**D**) Wild-type *SUP61* lacking an intron. (**E**) Wild-type *SUP61* with the presence of an intron (orange box). *SUP61* A1 and A2 refer to the two tRNA halves.

5.4.2. SUP61 *Variants Reveal 5' Flanking Sequence from* A. gossypii *is Responsible for Potential Precursor Accumulation*

An experiment was designed to identify the source of the additional band observed for SUP61 by generating a series of variants (**Figure 5.8**) containing alterations to the *SUP61* sequence. Each variant was integrated into the *HO* locus of strains lacking a wild-type copy of each. Northern blots using the SUP61 probe were then performed on biological

duplicates, before the membrane was stripped and re-probed with 5.8S rRNA as a normalising control.

The results (**Figure 5.9**) clearly show that the 5' flanking sequence from *Ashbya gossypii* is responsible for the additional band. All strains housing a wild-type 5' flanking sequence fail to produce the additional band. In contrast, all strains housing a synthetic 5' flanking sequence produce this additional band. The other variants show that strains housing a wild-type 3' flanking sequence and extra bases between the *SUP61* gene and 3' terminator all produced the extra band, indicating that the 3' region and strength of transcriptional terminator is not responsible. No additional band can be observed for the wild-type variant lacking an intron.

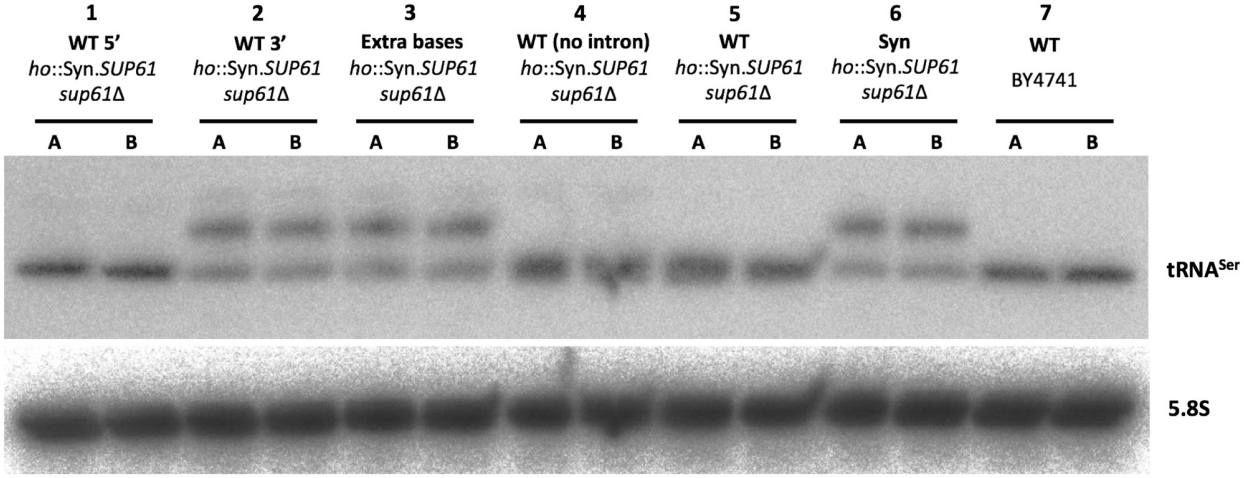


Figure 5.9: Northern blot results for each *SUP61* variant. The figure demonstrates that the 5' flanking sequence of *A. gossypii* is responsible for the additional band following Northern blot. An illustration of each construct may be found in **Figure 5.8**. Lanes A and B in the above figure indicate biological replicates (n = 2). A brief description of each variant by lane is summarised as follows:

1. *SUP61* with a wild-type 5' flanking sequence (variant A).
2. *SUP61* with a wild-type 3' flanking sequence (variant B).
3. *SUP61* containing extra bases before the 3' terminator sequence (variant C).
4. *SUP61* wild-type with no intron (variant D).

5. *SUP61* wild-type translocated to the *HO* locus (variant E).
6. A fully-synthetic variant of *SUP61* (positive control).
7. BY4741 wild-type (negative control).

5.4.3. *Discussion*

To identify the source of this additional SUP61 band, a series of SUP61 variants were constructed. The results (**Figure 5.9**) clearly shows that the 5' flanking sequence from *A. gossypii* was responsible. The significance of this finding is discussed in further detail in Section 6.1.3.

5.5. **Linearisation of the Neochromosome Constructed in BY4741**

The tRNA neochromosome was constructed in a stepwise manner from a smaller circular vector to the completed ~190kb circular 'megaplasmid'. However, a defining feature of true eukaryotic chromosomes is a linear structure replete with functional telomeres. Therefore, to be considered a true chromosome, the circular neochromosome must be linearised with the presence of functional telomeres on either end.

Mitchell and Boeke (2014) describes the use of a cassette designed to introduce synthetic telomere seed sequences (*TeSS*) into synthetic circular chromosomes. This section describes the introduction of this cassette into the tRNA neochromosome constructed in BY4741. BY4741 was initially identified due to the challenges of integrating the telomerator cassette into the triple-synthetic-chromosome background strain (described in further detail in Section 5.6).

As the *URA3* marker was required to integrate the telomerator cassette, the initial part of this section describes its removal on 5-FOA followed by linearisation using the telomerator, and visualisation using PFGE.

5.5.1. Experimental Approach

5.5.1.1. *URA3* 'Pop-out' from the tRNA Neochromosome

The circular neochromosome was constructed by repeatedly swapping the *LEU2* and *URA3* selective markers (with a static *HIS3* marker). As neochromosome linearisation required a *URA3* marker for integration of the telomerator cassette, the *URA3* marker was removed following the final round of inchworming by introducing a second UHR 'bridge' sequence in the final inchworm vector. This second UHR was designed to flank the *URA3* marker (**Figure 5.10**) and act as a region of homology to simply 'pop-out' the *URA3* marker on media containing the drug 5-FOA (5-fluoroorotic acid). The *URA3* gene converts 5-FOA into the toxic 5-fluorouracil – only cells that remove functionality of the *URA3* gene will survive. Neochromosome isolates that produced growth on SC-His + 5-FOA were verified for loss of the *URA3* marker by using a series of primers designed to amplify across the *URA3* junction region (these primers are indicated on **Figure 5.10**).

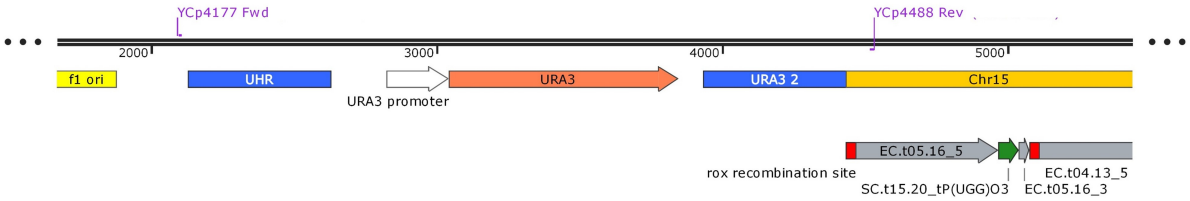


Figure 5.10: Graphical representation of the strategy used to remove the *URA3* gene on the neochromosome. A second 500 bp UHR region (blue box on the right) was incorporated into the final inchworm vector to flank the *URA3* selective marker (orange arrow; centre), facilitating its subsequent removal *via* homologous recombination on plates containing 5-FOA. The YCp4177 and YCp4488 primers on the above figure were used to screen for subsequent removal of the *URA3* marker, producing an amplicon of 637 bp for successful removal of the URA3 marker (and an amplicon of 2,443 bp if the *URA3* marker is still present).

5.5.1.2. Neochromosome Linearisation Using the Telomerator

The telomerator (**Figure 5.11**) is a cassette designed by Mitchell and Boeke (2014) and is designed to convert circular DNA vectors into linear chromosomes with functional telomeres. This cassette contains an intron inside a *URA3* auxotrophic marker – all intervening sequences are simply spliced away, thus preserving *URA3* functionality. Inside the intron lie telomere seed sequences (*TeSS*) flanking an 18 bp I-SceI recognition site that form the core of the telomerator cassette. Expressing the I-SceI homing endonuclease from a second episomal vector converts the circular structure into its linear form, releasing telomere seed sequences that are acted upon by telomerase to produce functional telomeres. Linearisation completely disrupts the coding frame of the *URA3* marker, and so provides selection.

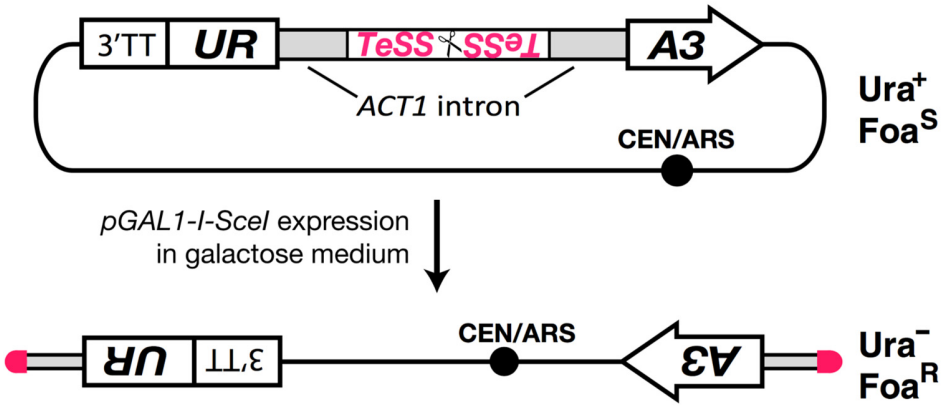


Figure 5.11: Representative diagram describing functionality of the telomerator cassette. The black line on the above image may also represent the body of the tRNA neochromosome. The presence of the *ACT1* artificial intron preserves *URA3* functionality – induction of a gene encoding I-SceI in galactose allows the incision (depicted by scissors) of the telomerator cassette, subsequently releasing telomere seed sequences (*TeSS*). The *URA3* marker acts as a form of selection: successful linearisation will remove functionality of the *URA3* marker and allow growth on media containing 5-FOA. The above figure was adapted from Mitchell and Boeke (2014).

Integration of the telomerator cassette into the tRNA neochromosome was undertaken by PCR-amplifying the telomerator cassette with primers containing 40 bp regions of homology to the site of integration. Three locations were identified for neochromosome linearisation (**Figure 5.13**): the first was adjacent to the centromere and the second two were located distal to the centromere. The second two sites were designed to replace the *Ter3* and *Ter2* cassettes, respectively, to ensure full replication of telomere arms. PCR fragments were transformed into strains housing the neochromosome, with successful integration of each telomerator cassette verified using colony PCR with primers designed to amplify across the junction regions of integration (data not shown – the PFGE results in **Figure 5.15** demonstrate that this approach was successful). Following neochromosome linearisation and verification (again using colony PCR), isolates housing the linearised neochromosome were subjected to pulsed-field gel electrophoresis. A flow chart summarising the primary steps involved in neochromosome linearisation may be found in **Figure 5.12**.

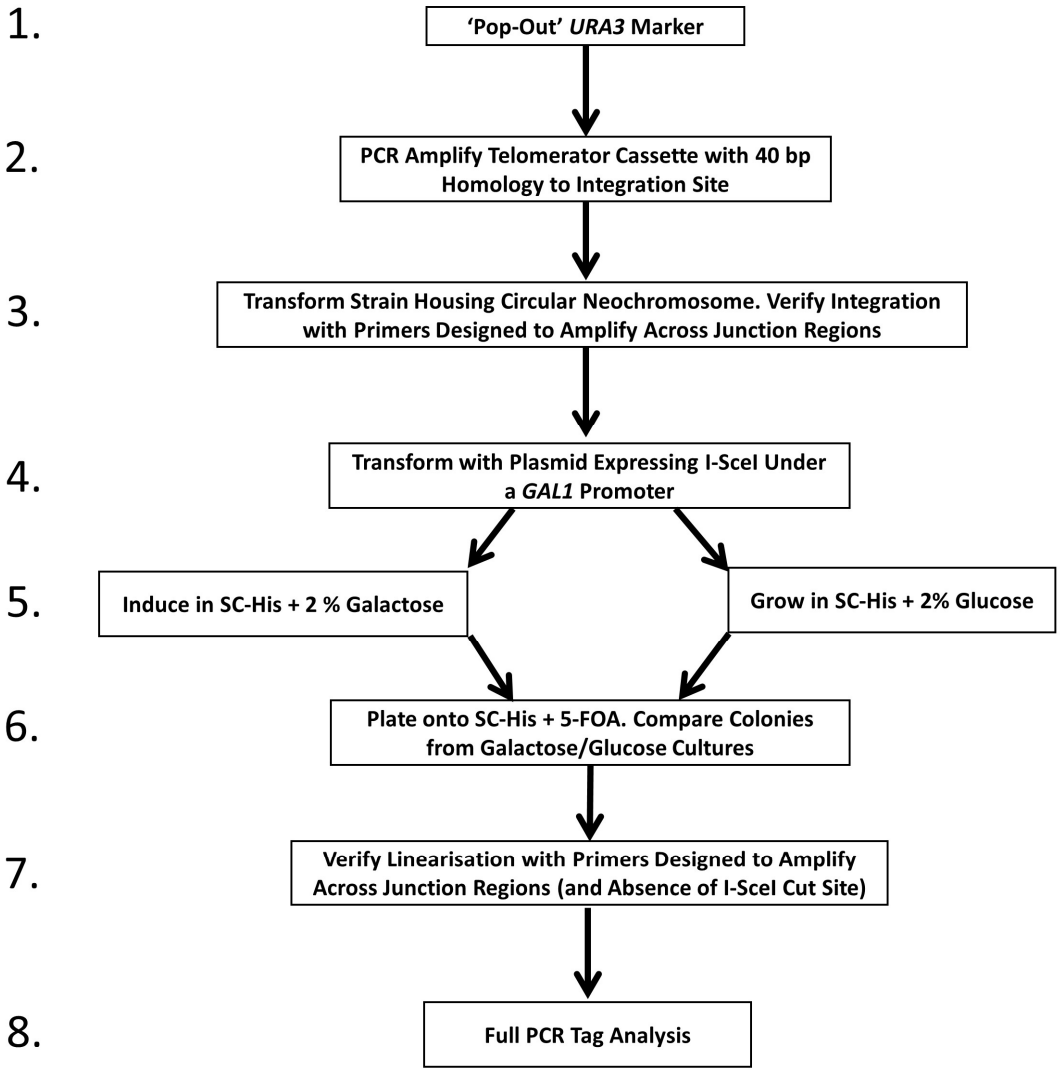


Figure 5.12: Flowchart summarising the primary steps involved in linearising the tRNA neochromosome. Part 6 is intended to demonstrate the efficiency of neochromosome linearisation by comparing the number of colonies induced in either 2% galactose or grown in 2% glucose. For part 7, successful cutting of the I-SceI target site may be inferred by the absence of a band on a PCR gel using primers designed to amplify across the junction region.

Figure 5.13: Diagram indicating the three sites of integration for the telomerator cassette. The telomerator cassette was amplified using PCR with homology arms for each integration site – the *URA3* marker provided a form of selection. Cassette A (**A** on the above diagram) is designed to be integrated proximal to the centromere, and cassettes B (**B** on the above diagram) and C (**C** on the above diagram) are designed to replace *Ter* cassette 3 and *Ter* cassette 2, respectively.

5.5.2. Results

5.5.2.1. PFGE of Neochromosome Constructed in BY4741 Reveals Unexpected Increase in Size

Linearisation of the neochromosome was first performed in the BY4741 strain background due to the problematic *URA3* remnant located in the *HO* locus of the triple-synthetic chromosome strain (described in further detail in Section 5.6). Direct visualisation of the linearised neochromosome was then undertaken using pulsed-field gel electrophoresis (**Figure 5.15**).

After plating for single colonies on SC-His + 5-FOA following induction in SC-His + 2% galactose, the results show that isolates housing the neochromosome linearised adjacent to the centromere (position A) produced very small colonies and one, single, large colony (**Figure 5.14A**). However, linearising the neochromosome at other positions distal to the centromere (position B and C) produced large colonies, indicating no adverse effect on colony size. In contrast, very few colonies were observed in isolates grown in SC-His + 2% glucose, suggesting a high efficiency of the telomerator system.

Pulsed-field gel electrophoresis on strains housing a linearised neochromosome at position B and position C unexpectedly revealed a complex duplication of the neochromosome structure, with a high degree of inter-colony heterogeneity (**Figure 5.15**). Two biological replicates of cassette B (lanes 2 and 6) produced a band at approximately 240 kb (50 kb larger than expected) and cassette C (lanes 4 and 7) produced bands at approximately 350 kb and 250 kb. The 350 kb band is notable for a near doubling in size of the neochromosome. The strain housing a circular neochromosome does not appear on the PFGE gel, likely due to the circular structure

becoming topologically trapped in the agarose fibres – it is known that circular chromosomes do not appear on PFGE gels (Hielm *et al.*, 1998, Mortimer *et al.*, 1990).

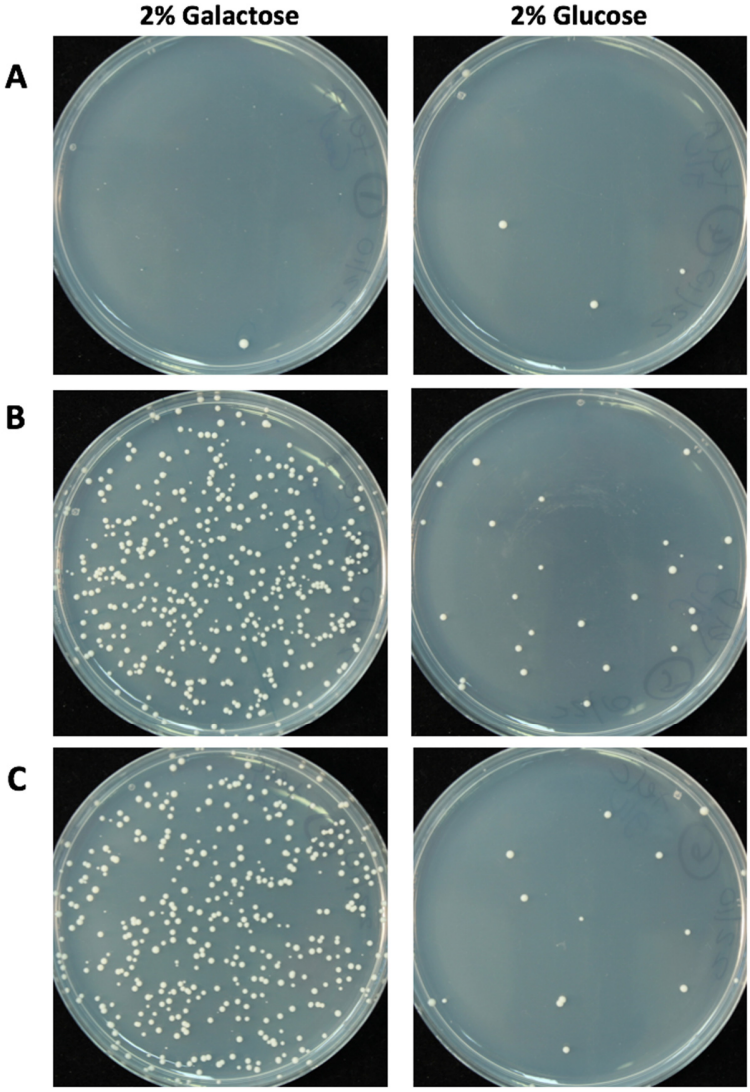


Figure 5.14: Yeast colonies on SC-His + 5-FOA plates following neochromosome linearisation. The plates above are intended to illustrate the efficiency of neochromosome linearisation by comparing the number of colonies on 5-FOA from cultures induced in SC-His + 2% galactose (left) and grown in SC-His + 2% glucose (right). Colonies on the above plates are inferred to have lost their *URA3* marker through either successful linearisation with I-SceI, or, alternatively, from loss of function of the *URA3* marker through other means (for example, a frame-shift mutation).

[Figure legend continued on following page]

Figure 5.14 (Continued). The latter explains the colonies on 5-FOA grown in 2% glucose. (**A**) is cassette A (linearised adjacent to the centromere). The following colonies were counted: 2% galactose: 1 colony, 2% glucose: 4 colonies. (**B**) is cassette B (linearised at *Ter* cassette 3). The following colonies were counted: 2% galactose: 544 colonies, 2% glucose: 26 colonies. (**C**) is cassette C (linearised at *Ter* cassette 2). The following colonies were counted: 2% galactose: 456 colonies, 2% glucose: 16 colonies.

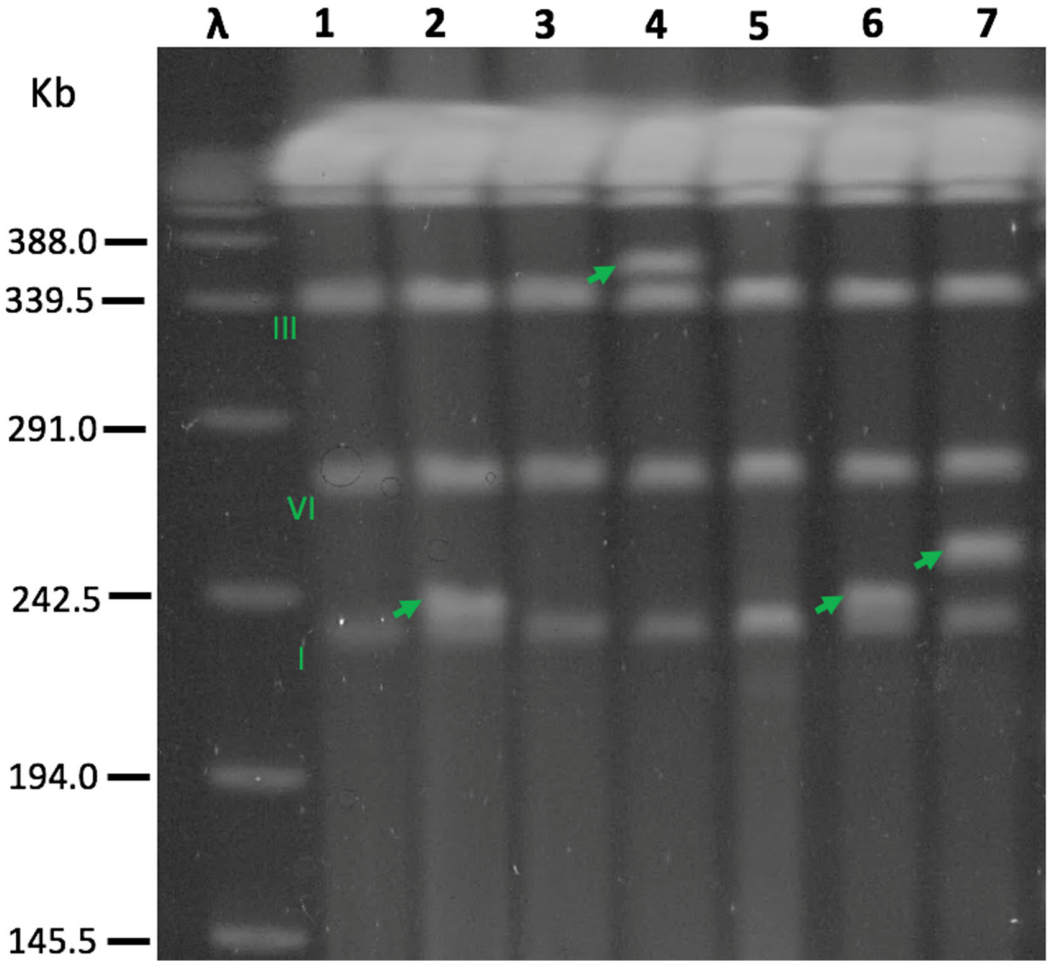


Figure 5.15: Pulsed-field gel pattern of BY4741 strains housing linear and circular variants of the tRNA neochromosome. The above gel image shows variants of the tRNA neochromosome linearised at positions B and C (as denoted in **Figure 5.13**) and shows an unexpected increase in size**.** Linearisation was performed on biological replicates (n = 2). The expected size of the tRNA neochromosome is ~187 kb. Wild-type yeast chromosomes are indicated by Roman numerals and bands corresponding to the neochromosome are indicated by green arrows.

[Figure legend continued on following page]

Figure 5.15 (Continued). The pulsed-field gel apparatus used switch times set to resolve lower molecular weight chromosome sizes, and so the larger *S. cerevisiae* chromosomes are not shown on the above gel. Lambda DNA was used as a molecular weight ladder. Lanes 1 and 3 are strains housing circular neochromosomes (n = 1), lanes 2 and 6 are replicate isolates (n = 2) of strains housing a neochromosome linearised at position B (both display a ~240 kb band), lanes 4 and 7 are replicate isolates (n = 2) of strains housing a neochromosome linearised at position C (displaying a band at ~350 kb and ~250 kb, respectively) and lane 5 is a wild-type control. Circular variants of the tRNA neochromosome are not expected to be visible on the above gel image.

5.5.2.2. *In vitro* Neochromosome Linearisation Reveals Chronology of Structure Variations

The neochromosome constructed in the BY4741 background was observed to display a significant increase in size. However, a heterogeneity in size was observed between isolates (**Figure 5.15**) and no proof could be deduced from sequencing results that the observed structural variations occurred prior to linearisation using the telomerator (**Figure 5.21**). To demonstrate that the observed increases in size occurred when the neochromosome existed in its circular form, an *in vitro* linearisation assay was performed. Agarose plugs containing the neochromosome with integrated telomerator cassettes at position A, B and C were subjected to a restriction-enzyme digest using I-SceI. Following restriction digest, PFGE was performed to directly visualise linearised variants of the neochromosome.

The results (**Figure 5.16**) demonstrate that the increase in neochromosome size occurred when the neochromosome existed in its circular form. *In vitro* linearisation of biological replicates housing the neochromosome at position B and position C produced approximate bands at ~240 kb and ~250 kb, respectively. However, linearisation of the neochromosome at position A revealed two faint bands at ~160 kb and ~265 kb. It is possible these two bands are caused by a heterogeneity of neochromosome size in the

cell population. No band corresponding to the neochromosome can be observed for circular controls, showing that the I-SceI restriction enzyme successfully linearised the neochromosome. It should be noted that the approximate doubling in size (**Figure 5.15**, well 4) for the neochromosome linearised at position C was not observed, suggesting that neochromosome concatenation may have occurred after *in vivo* linearisation.

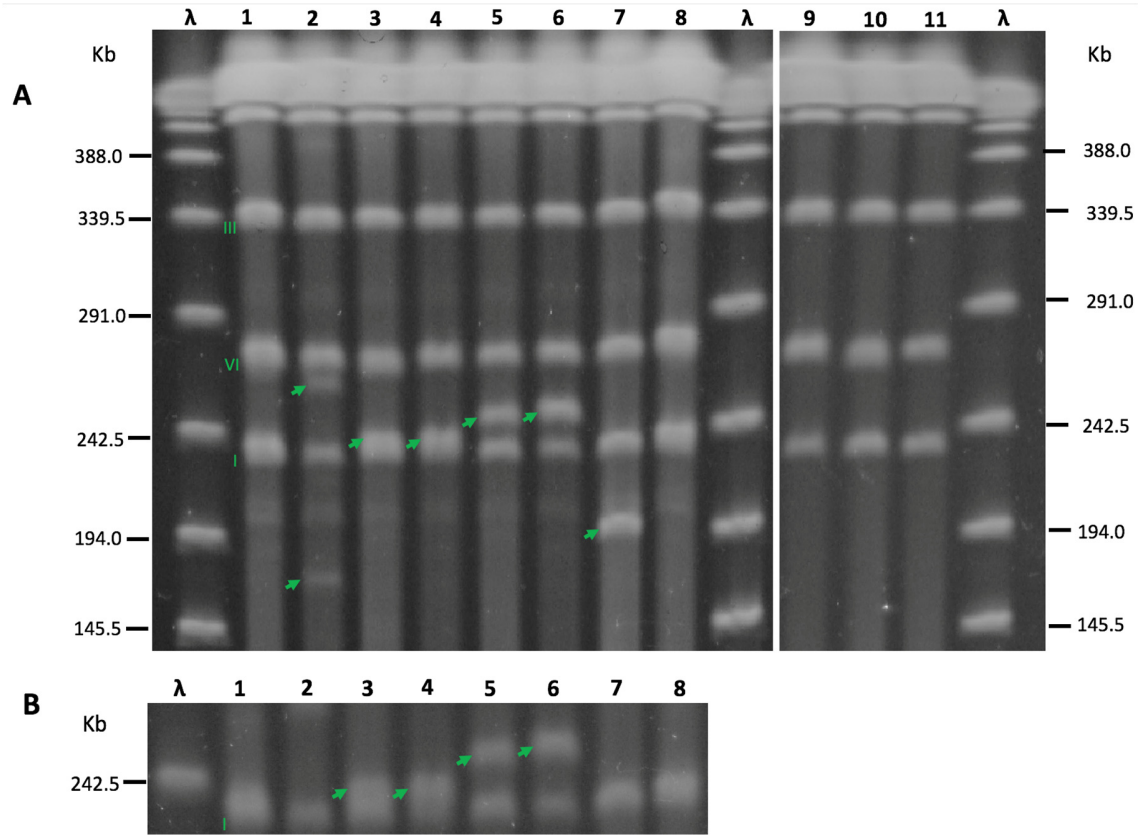


Figure 5.16: Pulsed-field gel pattern of the tRNA neochromosome subjected to *in vitro* restriction enzyme digest with I-SceI. The above gel image shows variants of the tRNA neochromosome linearised at positions A, B and C (as denoted in **Figure 5.13)** using the I-SceI restriction enzyme. The expected size of the tRNA neochromosome is ~187 kb. Wild-type yeast chromosomes are indicated by Roman numerals and bands corresponding to the neochromosome are indicated by green arrows. Lambda DNA was used as a molecular weight ladder. (**A**) is the full PFGE gel and (**B**) is a cropped image of the same gel run for an additional 8 hours to improve separation of the neochromosome bands in lane 3 and 4. Lanes 1 and 8 are biological replicates (n = 2) of wild-type (BY4741 + pRS413) controls, lane 2 is a strain housing the neochromosome with the telomerator at position A (n = 1), lanes 3 and 4 are biological replicates (n = 2) of the neochromosome with the telomerator at position B, lanes 5 and 6 are of the telomerator at position C (n = 2) and lane 7 is a positive control (tRNA neochromosome constructed in the triple-synthetic-chromosome background housing the telomerator cassette at position B). Lanes 9, 10 and 11 are circular negative controls (n = 1) of the tRNA neochromosome housing the telomerator cassette at position A (lane 9), position B (lane 10) and position C (lane 11). Only lanes 1 to 8 were subjected to restriction digest using I-SceI. Circular variants of the tRNA neochromosome are not expected to be visible on the above gel image.

5.5.3. Discussion

Linearisation of the neochromosome adjacent to the centromere produced exceptionally small colonies when plated onto 5-FOA, compared to larger colonies when linearised distal to the centromere (**Figure 5.14**). There are several potential sources for this phenomenon. One possibility includes incomplete replication caused by disrupting the normal path of the replication fork *i.e.* the *Ter1* replication termination cassette should theoretically block replication of a region of DNA between the *Ter1* site and telomere arm following linearisation. Other potential sources of this small colony phenotype include a destabilising effect of the telomeres on centromere function, a silencing effect on the function of the *HIS3* marker or, alternatively, an imbalance in chromosome size due to one telomere arm being significantly longer than the other.

Linearisation using the telomerator cassette and subsequent visualisation of the neochromosome constructed in the BY4741 strain background revealed a large and unexpected increase in size (**Figure 5.15**). A subsequent *in vitro* linearisation performed using the I-SceI restriction enzyme (**Figure 5.16**) revealed that the unexpected increase in size occurred when the neochromosome existed in its circular form. This demonstrates that linearisation of the tRNA neochromosome does not lead to concatenation between two neochromosome pairs.

5.6. **Linearisation of the Neochromosome Constructed in Syn III/VI/IXR**

The neochromosome was constructed in two independent *S. cerevisiae* strain backgrounds: BY4741 and a strain housing a triple-synthetic-chromosome (SynIII, SynVI and SynIXR). Following PFGE, the neochromosome constructed in the BY4741 strain background was unexpectedly observed to house a large, complex internal duplication

(**Figure 5.15**). To circumvent this issue, it was necessary to linearise the neochromosome in the triple-synthetic-chromosome strain.

However, an additional challenge was observed due to the discovery of a non-functional *ura3* marker containing point mutations in the *HO* locus of this strain, rendering integration of the *URA3*-based telomerator near-impossible. This *ura3* marker was caused by an error during construction of SynIII (Mitchell *et al.*, 2017). For successful integration of the telomerator cassette, the initial part of this section describes the chemical extraction, purification and transfer of the circular neochromosome into a new host using spheroplast transformation. The latter part of this chapter describes the integration of the telomerator cassette, subsequent linearisation and visualisation using PFGE. Finally, a plate reader growth assay was performed to compare both linear and circular variants.

5.6.1. Experimental Approach

5.6.1.1. Neochromosome Extraction and Transfer

The circular structure of the neochromosome presents the opportunity for its separation from linear genomic DNA. Devenish and Newlon (1982) describe a method using the properties of high pH to denature linear DNA, followed by a high-salt-phenol extraction to separate a ring version of yeast chromosome III from the other linear chromosomes. This method was attempted, although the purity of recovered DNA was low and large quantities of unwanted RNA was observed (data not shown).

An updated protocol applicable to synthetic biology was devised by Noskov *et al.* (2011). This method has several important differences to Devenish and Newlon (1982), including an increased volume of yeast culture, increased pH for spheroplast lysis followed by a

phenol:chloroform:isoamyl extraction and isopropanol precipitation. RNA is then removed from the DNA with RNase A followed by further purification, including exonuclease digestion, with the Large Construct Kit (Qiagen). Neochromosome DNA was purified using this method and subsequently transformed into yeast using the spheroplast transformation method.

5.6.2. Results

5.6.2.1. Chemical Extraction and Transfer Facilitates Linearisation of the Neochromosome Constructed in Syn III/VI/IXR

The neochromosome constructed in the BY4741 background displayed an unexpected increase in size (**Figure 5.15**). To integrate the telomerator into the neochromosome constructed in the triple-synthetic-chromosome background, it was first necessary to chemically extract and transfer this neochromosome back into BY4741. Neochromosome extraction was undertaken, with modifications, according to Noskov *et al.* (2011).

Figure 5.17A shows representative yeast plates comparing the number of colonies obtained from lithium acetate transformation and spheroplast transformation of purified neochromosome DNA. Approximately ten times as many colonies were observed following the spheroplast transformation method. All colonies produced from the lithium acetate method tested negative for the presence of the Chr7 and Chr12 tRNA arrays, suggesting a low or no efficiency of transfer. **Figure 5.17B** is an indicative colony PCR performed following spheroplast transformation to screen for one tRNA cassette on the Chr16 tRNA array. Eight of the twelve colonies tested positive for the presence of the Chr16 tRNA array, with two of these colonies further verified for the presence of an intact neochromosome (confirmed using PFGE of these isolates in Section 5.6.2.2.). The data suggests that the chemical neochromosome extraction and spheroplast transfer method

is highly efficient (up to 66% of isolates contain a successfully transferred neochromosome).

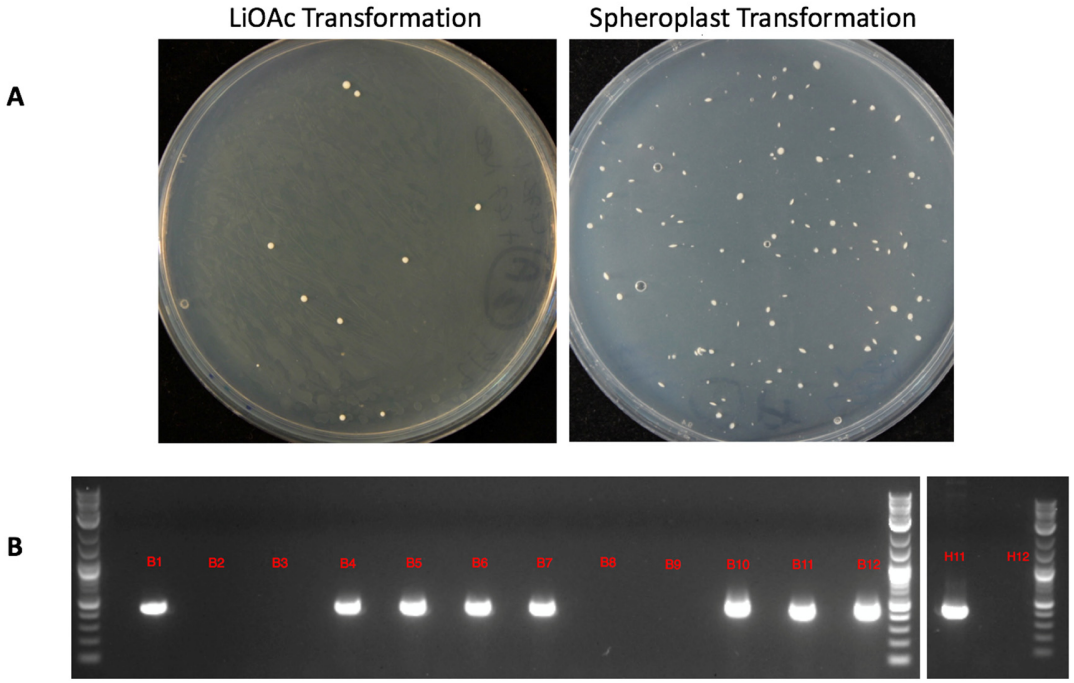


Figure 5.17: Representative results comparing lithium acetate transformation with spheroplast transformation using purified circular neochromosome DNA. The above figure illustrates the efficiency of neochromosome transformation by comparing the number of colonies following the lithium acetate transformation method and that of spheroplast transformation. (**A**) The plate on the left shows yeast colonies on SC-His following transformation using the lithium acetate method. The plate on the right displays embed yeast colonies in SC-His (plus sorbitol) following spheroplast transformation. A total of 10 colonies can be observed from LiOAc transformation and 105 colonies following spheroplast transformation. (**B**) Representative colony PCR to verify the presence of the SC.t16.12 tRNA cassette on the Chr16 tRNA array on isolates recovered following spheroplast transformation (n = 12). The expected amplicon size is 447 bp. Well H11 is the positive control and well H12 is the negative control. A representative 2-log ladder with annotated DNA fragment sizes may be found in Figure 2.1

5.6.2.2. PFGE of Neochromosome Constructed in Syn III/VI/IXR Reveals Correct Approximate Size

Following neochromosome extraction and transfer to BY4741, the telomerator cassette was then again integrated into two regions distal to the centromere (cassette B and C) prior to linearisation and PFGE. This was performed according to the method described in Section 5.5.1.2. The results (**Figure 5.19**) show that the linearised neochromosome constructed in the triple-synthetic-chromosome strain background produce the expected approximate size (~190 kb), although a slight increase in size can be observed. This shows that the large increase in neochromosome size is restricted to the version constructed in BY4741.

Figure 5.18: Diagram indicating the three sites of integration for the telomerator cassette. The above figure is intended to illustrate the locations of neochromosome linearisation described in **Figure 5.19**. The telomerator cassette was amplified using PCR with homology arms for each integration site – the *URA3* marker provided a form of selection. Cassette A (**A** on the above diagram) is designed to be integrated proximal to the centromere, and cassettes B (**B** on the above diagram) and C (**C** on the above diagram) are designed to replace *Ter* cassette 3 and *Ter* cassette 2, respectively.

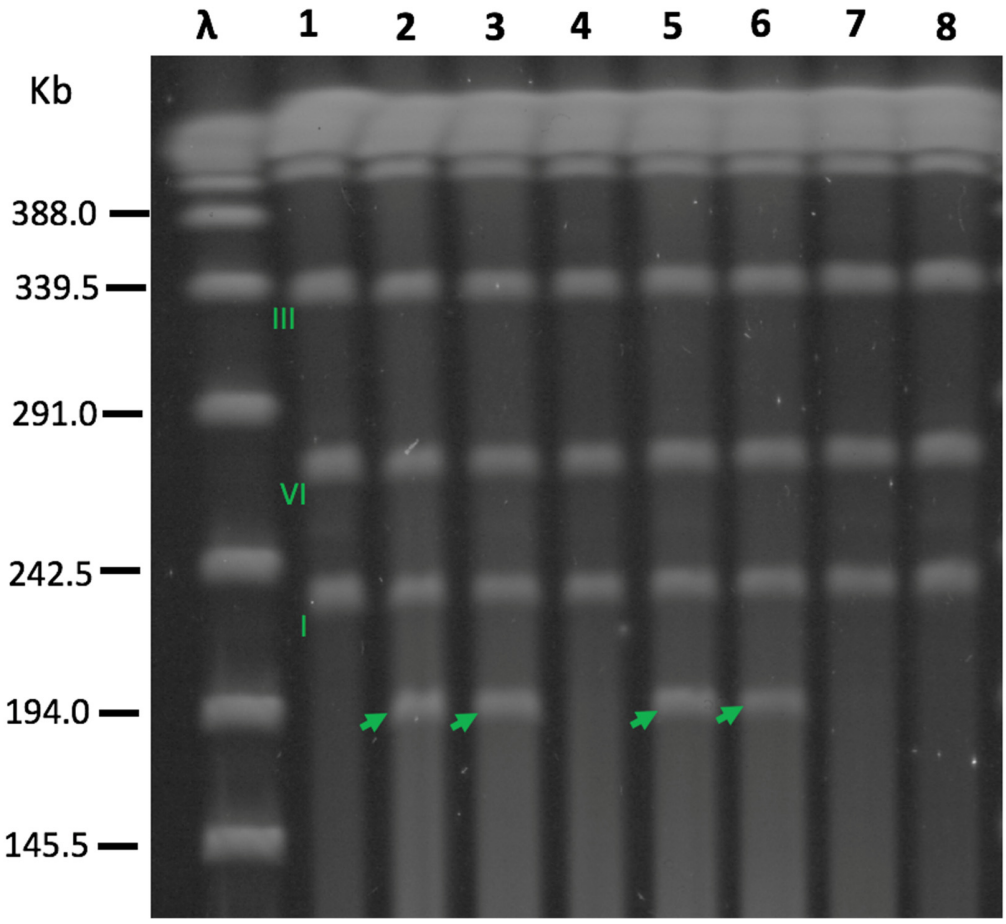


Figure 5.19: Pulsed-field gel pattern of the neochromosome constructed in the triple-synthetic-chromosome background and subsequently transferred back to BY4741. The above gel image shows variants of the tRNA neochromosome linearised at positions B and C (as denoted in **Figure 5.18**), as biological replicates (n = 2). The expected size of the tRNA neochromosome is ~187 kb. Wild-type yeast chromosomes are indicated by Roman numerals and bands corresponding to the neochromosome are indicated by green arrows. Lambda DNA was used as a molecular weight ladder. Lanes 2 and 3 are isolates of strains housing a neochromosome linearised at position B (n = 2), lanes 5 and 6 are isolates of strains housing a neochromosome linearised at position C (n = 2). Lane 1 and lane 8 are wild-type controls and lanes 4 and 7 are strains housing circular neochromosomes. Circular variants of the tRNA neochromosome are not expected to be visible on the above gel image.

5.6.2.3. Plate Reader Growth Assay of the tRNA Neochromosome Reveals Slow Growth Phenotype

A plate reader assay was used to determine the growth profile of strains housing variants of the neochromosome constructed in the triple-synthetic-chromosome background and transferred back to BY4741. Neochromosome isolates linearised at position B, position C and a circular parental variant were compared to BY4741 + pRS413 wild-type control.

All isolates were inoculated into SC-His (2% glucose) as biological triplicates and incubated with rotation at 30^{o}C overnight. The following morning, cultures were re-inoculated into fresh media to a target OD_{600nm} of 0.1 and run on a plate reader using standard settings. The results in **Figure 5.20** show a slow growth phenotype for all strains housing a tRNA neochromosome compared to BY4741 + pRS413. No significant difference can be observed for strains housing linear and circular forms of the neochromosome. Approximate doubling times were calculated to be 2.95 hours for BY4741 + pRS413 and between 3.45 and 3.68 hours for strains housing the tRNA neochromosome, representing a difference in logarithmic-phase growth of approximately 17%.

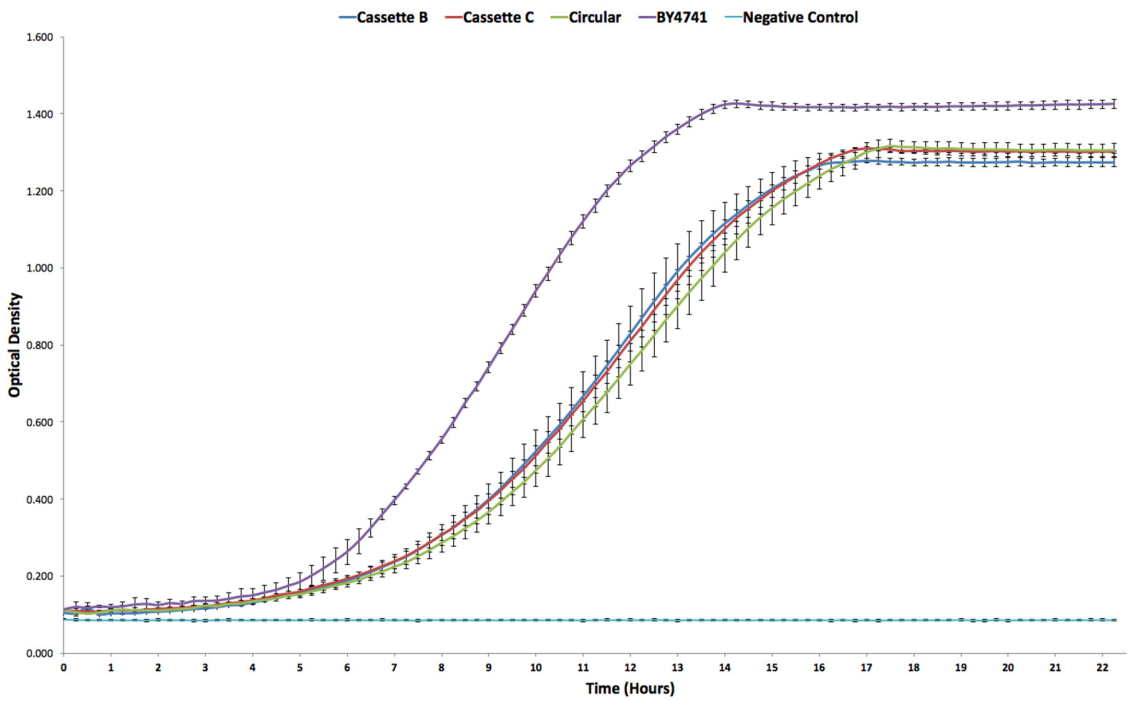


Figure 5.20: Plate reader growth assay of BY4741 strains housing linear and circular variants of the tRNA neochromosome. The above chart demonstrates a slow-growth phenotype in strains housing the tRNA neochromosome. Each growth curve on the above chart are means of biological replicates (n = 3), and error bars are the standard deviation of these values. Cassette B (blue line) is the tRNA neochromosome linearised at position B, Cassette C (red line) is linearised at position C, Circular (green line) is the circular version of the neochromosome and BY4741 (purple line) is BY4741 + pRS413. The negative control (turquoise line) is SC-His liquid media without inoculum.

5.6.3. Discussion

The *HO* locus of the triple-synthetic-chromosome strain contains a non-functional, remnant *ura3* marker caused by an error during construction of SynIII (Mitchell *et al.*, 2017). Previous attempts at integrating the *URA3*-based telomerator cassette into this strain induced repair of *ura3*, resulting in a failure to identify any isolate containing the successfully integrated telomerator cassette (data not shown).

Although the *HO* locus may have simply been repaired, the chemical extraction and transfer method described previously not only circumvented the challenge of telomerator integration, but also demonstrated that the neochromosome is indeed a physical object. Additionally, this method provides a tool for the Sc2.0 consortium to transfer the neochromosome into any strain required.

Spheroplast transformation of purified neochromosome DNA was observed to be critical (**Figure 5.17**). Purified DNA was transformed directly into BY4741 using the lithium acetate transformation method but transformation efficiency was low and all isolates tested negative for the presence of an intact neochromosome. The large size of the neochromosome DNA, coupled with the 'insult' of using relatively toxic lithium acetate and DMSO, are known to be unsuitable for the transformation of chromosome-size DNA fragments, such as yeast artificial chromosomes (Heale *et al.*, 1994). Spheroplast transformation eliminates the presence of a cell wall that would otherwise act as a barrier to large DNA fragments of between 100 kb to 1 Mb in size (Kawai *et al.*, 2010). Transformation of yeast using spheroplast transformation was observed to be highly efficient – 8 out of 12 isolates potentially tested positive for presence of the neochromosome.

To support the model(s) of neochromosome instability proposed in Section 6.1.5, it was necessary to prove that *in vivo* neochromosome linearisation was not directly responsible for the observed structural variations. An *in vitro* restriction enzyme digest using I-SceI revealed that the structural variations occurred when the neochromosome existed in its circular form (**Figure 5.16**).

Finally, the plate reader growth assay performed on strains housing the neochromosome revealed a slow growth phenotype compared to BY4741 + pRS413 (**Figure 5.20**). No significant differences were observed for linearised neochromosome isolates or the

circular forms, suggesting that linearisation does not confer a significant growth advantage. Potential sources of any slow growth phenotype are numerous, and briefly described in Section 6.1.3.

5.7. **Deep Sequencing of the tRNA Neochromosome**

PFGE of the neochromosome constructed in BY4741 revealed an unexpected increase in size. To identify the source of this increase and verify the integrity of the neochromosome constructed in the triple-synthetic-chromosome strain, deep sequencing and analysis was performed by Beijing Genomics Institute (Shenzhen, China).

5.7.1. Results

5.7.1.1. Deep Sequencing of the tRNA Neochromosome Constructed in BY4741

Three of the BY4741 strains displaying a significant increase in neochromosome size in **Figure 5.15** were sequenced: these include lane 2 (linearised at position B; ~240 kb band following PFGE), lane 4 (linearised at position C; ~350 kb band) and lane 7 (linearised at position C; ~250 kb band).

Sequencing results revealed that the unexpected increase in neochromosome size in BY4741 was caused by a complex internal duplication (**Figure 5.20**). In all three cases, structural variations were observed to have occurred between *rox* recombination sites. It should be noted, however, that sequence read depth was only 10X, and may account for the discrepancies in neochromosome size from sequencing results and PFGE. Due to the low sequence read depth, these sequencing results may be considered preliminary, although do provide enough read quality to indicate the location of structural variations.

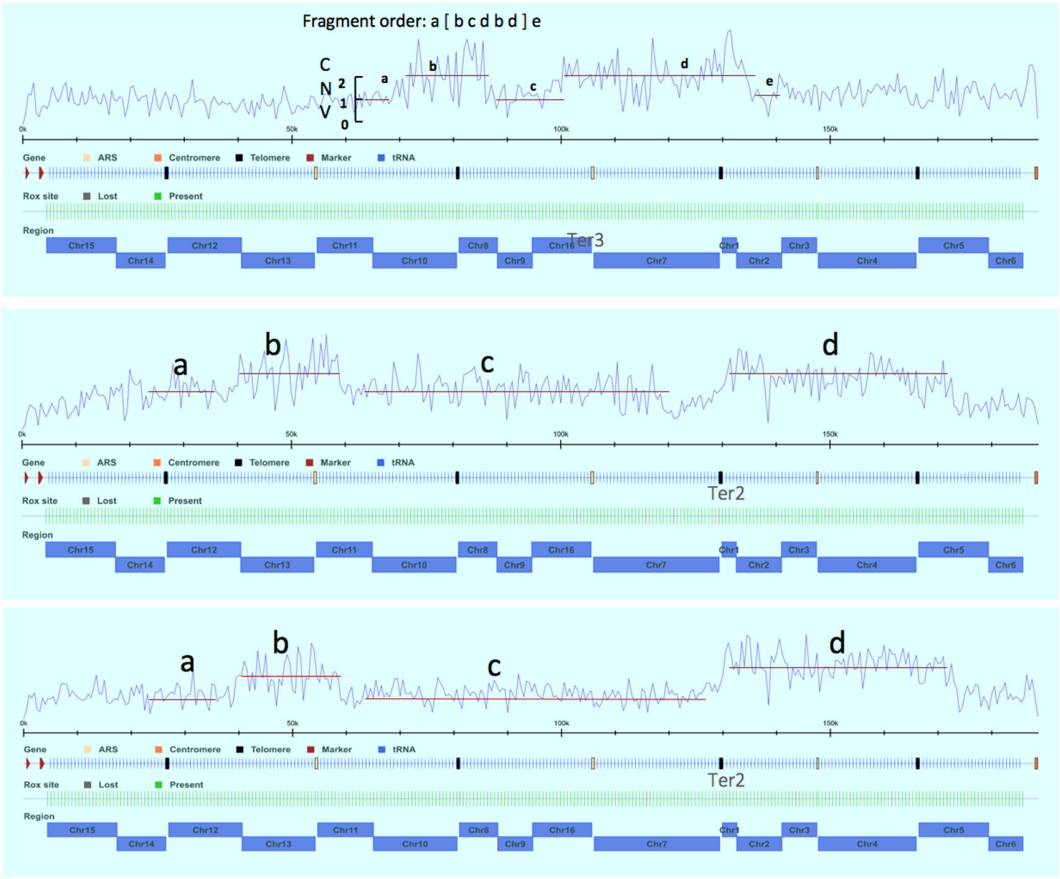


Figure 5.21: Preliminary deep-sequencing analysis of three BY4741 isolates following neochromosome linearisation. The above preliminary sequencing results (n = 1) show a doubling in sequencing reads at separate points on the tRNA neochromosome, indicating structural variations. The above sequenced isolates all displayed a significant increase in size following PFGE. In the above figure, CNV indicate the copy number variations of each sequenced strain, with fragment order (a, b, c, d and e) describing the nature of duplications observed. In all cases, fragment b and fragment d displayed a doubling of sequencing read depth, indicating a duplication of these regions. Sequencing read depth is 10X. (**A**) neochromosome linearised at position B producing a ~240 kb PFGE band. (**B**) neochromosome linearised at position C producing a ~350 kb band. (**C**) neochromosome linearised at position C producing a ~250 kb band.

5.7.1.2. Deep Sequencing of the tRNA Neochromosome Constructed in Syn III/VI/IXR

Sequencing of the linearised neochromosome constructed in the triple-synthetic-chromosome strain backround (**Figure 5.19**) was also performed following extraction and transfer back to BY4741. Sequencing results (**Figure 5.22**) show a largely intact neochromosome, with the exception of a structural variation. The same structural variations were observed for the two neochromosome variants linearised at position B (**Figure 5.22A**) and position C (**Figure 5.22B**).

A duplication can be observed for part of the Chr13 tRNA array, in addition to a large inversion between the Chr13, Chr11 and central region of the Chr10 tRNA array. Sequencing results also reveal the small deletion of two tRNA cassettes – these include the *tR(UCU)J2* (Arg) and *tV(AAC)J* (Val) tRNA genes, both existing as 11 copies and 13 copies in the *S. cerevisiae* genome, respectively. The loss of these two tRNA cassettes may also be inferred by their absence following PCR tag analysis (**Figure 5.22C**).

An inversion was also unexpectedly observed for the CEN/ARS region, although this is unlikely to affect centromere function. A sequencing 'spike' was observed near the Chr1 tRNA array in the strain linearised at position B – this was found to be a sequencing artefact caused by high homology between the *Ter2* replication termination cassette and the rDNA locus of the *S. cerevisiae* genome.

In all examples, structural variations were observed to have occurred between *Rox* recombination sites. **Figure 5.23** is a representative diagram of the neochromosome containing annotated structure variations. **Table 5.2** is a summary of variations identified following deep sequencing of the neochromosome constructed in the triple-synthetic-chromosome background. With the exception of structural variations identified in **Figure 5.22**, sequencing revealed a total of 11 variations, including deletions, single-nucleotide

variations and insertions. Alterations to the *chrF-444* origin of replication were observed following Sanger sequencing of the original plasmid, and were not observed to affect the function of this origin (**Figure 5.2**). The other alterations are considered minor, and not likely to affect neochromosome function.

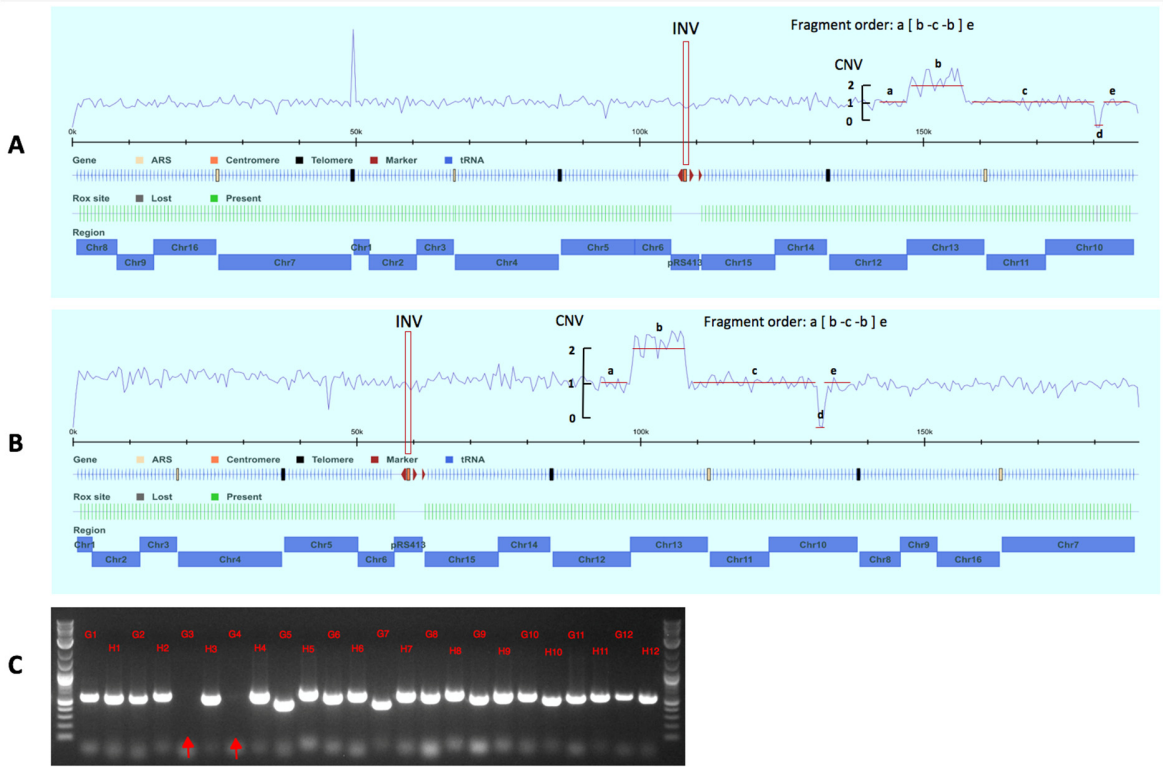


Figure 5.22: Deep-sequencing analysis of the tRNA neochromosome constructed in the triple-synthetic-chromosome strain background. The above sequencing results show a doubling in sequencing reads at separate points on the tRNA neochromosome, indicating structural variations. The above isolates (n = 1) housed the neochromosome linearised position B (**A**) and position C (**B**) following extraction and transfer back to BY4741. The inversion of the region around the centromere is indicated by INV, as are copy number variations (CNV). The fragment order (a, b, c, d and e) describe the order of the Chr13 tRNA array duplication and translocation events into the Chr10 tRNA array. Corresponding tRNA arrays are indicated by blue blocks on the bottom of each diagram. Sequencing read depth is between 81 X and 92 X. (**C**) A partial PCR-tag assay (plate 2). The absence of two tRNA cassettes corresponding to the deletion are indicated by two red arrows (wells G3 and G4). The corresponding tRNA genes for each well may be found in Appendix III, **Table II**. A representative 2-log ladder with annotated DNA fragment sizes may be found in **Figure 2.1**

Table 5.2: Summary of variations observed in the neochromosome sequence following sequencing. The following table demonstrates the low number of sequence alterations following neochromosome construction. The following variations occurred for the neochromosome constructed in the triple-synthetic-chromosome strain background. Both neochromosomes linearised at position B and position C contained the same variations. 'SNV' refer to single nucleotide variations, 'INS' refers to insertions 'DEL' refers to deletions, 'reference' describes the predicted nucleotide(s) and 'alteration' describes the observed nucleotide change.

Position	Reference	Alteration	Type	Region Affected
67220	T	C	SNV	ARS *chrF-444*
67342	T	TA	INS	ARS *chrF-444*
67481	T	G	SNV	ARS *chrF-444*
67482	A	C	SNV	ARS *chrF-444*
86206	TA	T	DEL	*Rox* recombination site
88684	C	A	SNV	SC.t05.04 5' flanking sequence
99133	G	C	SNV	SfiI restriction site
99742	G	C	SNV	SfiI restriction site
101034	T	A	SNV	SfiI restriction site
102290	T	A	SNV	SfiI restriction site
147093	TA	T	DEL	*Rox* recombination site

5.7.2. Discussion

Deep sequencing of the tRNA neochromosome revealed structural variations in the variants constructed in both BY4741 and the triple-synthetic chromosome. Due to the low sequencing read depth of the neochromosome constructed in the BY4741 background, these results may be considered preliminary. However, they do provide an indication of the structural variations following PFGE (**Figure 5.15**). Although the neochromosome constructed in the triple-synthetic-chromosome produced a near-correct size following PFGE (**Figure 5.19**), deep sequencing revealed that this strain too

contained a structural variation. This included a duplication of part of the Chr13 tRNA array and an inversion of a ~32.8 kb region (**Figure 5.22**). This duplication and inversion appears to be responsible for the loss of two tRNA genes, although their loss is not a significant concern due to their high-copy number in the *S. cerevisiae* genome.

With the exception of structural variations, deep sequencing revealed a low number of DNA sequence errors following neochromosome construction (**Table 5.2**). A total of 11 minor variations were observed in the neochromosome sequence, indicating that the neochromosome construction process is largely seamless. Finally, it's notable that deep sequencing revealed that structural variations are caused by recombination between *rox* recombination sites.

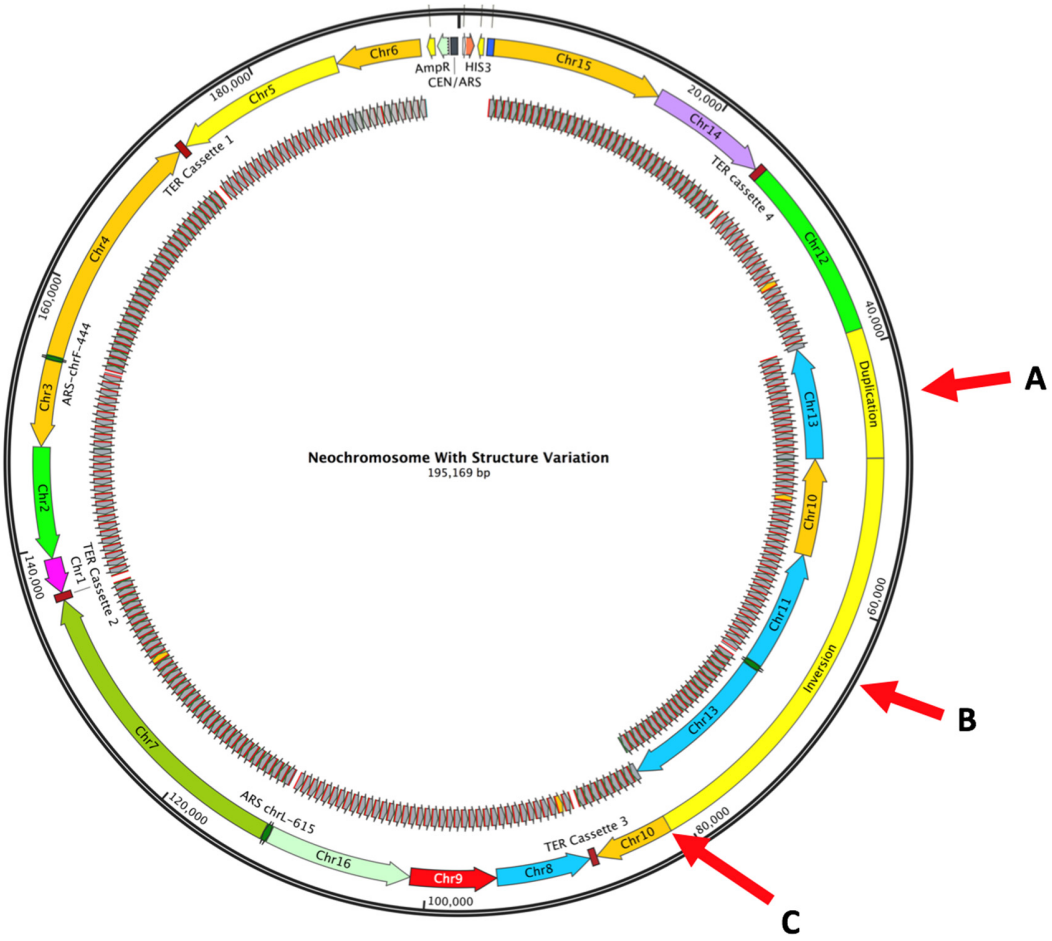


Figure 5.23: Representative graphical map of tRNA neochromosome constructed in the triple-synthetic-chromosome with annotated structure variations. The above structural variations were inferred following deep sequencing (**Figure 5.22**), and indicate an increase in neochromosome size to 195 kb. On the above image, (**A**) denotes the duplicated section of the Chr13 tRNA array (yellow 'duplication' box), (**B**) indicates the inverted region containing the Chr13, Chr11 and Chr10 tRNA arrays (yellow 'inversion' box) and (**C**) indicates the site of deletion of two tRNA cassettes.

5.8. General Discussion on Neochromosome Characterisation

This chapter focused on two facets of characterisation, including the characterisation of elements associated with neochromosome function and physical manipulation of the tRNA neochromosome structure. The former, while limited in scope, demonstrates that

elements of the neochromosome do function, although an additional band may be observed for the SUP61 tRNA following Northern blot. The latter, including neochromosome linearisation and chemical extraction and transfer, demonstrates that the neochromosome is an intact, physical object. Sequencing results, however, show structural variations for the tRNA neochromosome constructed in BY4741 and the triple-synthetic-chromosome.

Elements associated with neochromosome replication include orthogonal origins of replication and replication termination sites. Simple assays were used to characterise these elements, and revealed that they do function, although not in a strictly quantitative manner. Due to a lack of replicates, these results may be considered preliminary, although more powerful studies into replication in the context of the neochromosome using deep-sequencing may be undertaken in future (Muller et al., 2014).

The results in Section 5.3 revealed that three, synthetic, single-copy tRNA genes (*SUP61*, *TRT2* and *TRR4*) can support viability in BY4741 in the absence of their wild-type counterparts. Notably, the synthetic SUP61 variant produced an additional band following Northern blot. To elucidate the source of this additional band, a series of *SUP61* variants were constructed and clearly identify the synthetic 5' flanking sequence as the element responsible. The significance of these findings is discussed further in Chapter 6.

To be considered a true eukaryotic chromosome, it was necessary to linearise the circular neochromosome and introduce functional telomeres. Subsequent PFGE of the variant constructed in BY4741 revealed a significant increase in size. Although a non-functional *ura3* marker was identified in the triple-synthetic-chromosome variant hindered integration of the telomerator cassette, the chemical extraction and transfer method back to BY4741 circumvented this issue. The chemical extraction and transfer method

also demonstrates the feasibility of transferring the tRNA neochromosome into separate strain backgrounds – a critical requirement for the Sc2.0 consortium.

A plate reader growth assay revealed that linearisation of the tRNA neochromosome confers no overall advantage when compared to the circular variants. An increase in doubling time also suggests that the presence of the tRNA neochromosome produces a metabolic burden on the cell.

Subsequent deep sequencing revealed that both versions of the neochromosome constructed in BY4741 and the triple-synthetic-chromosome contain structural variations. Speculation on the source of these structural variations is outlined in Chapter 6. Finally, it's notable that construction of the tRNA neochromosome is largely seamless, with few nucleotide variations observed following deep sequencing (**Table 5.2**). This demonstrates the power of the inchworming method and *in vivo* homologous recombination in yeast.

Chapter 6: General Discussion and Future Scope

6.1. Discussion – Summary and Significance of Work

The tRNA neochromosome was successfully constructed from its constituent tRNA arrays, differing only to the original design by some structural variations. Overall, this doctoral study applied new approaches to design, construct and characterise a tRNA neochromosome in the yeast model, *Saccharomyces cerevisiae*. This project has also contributed an important component for the Sc2.0 consortium.

6.1.1. Engineering Approaches Rationalise the Complexities of Neochromosome Design

This project presented an opportunity to design of one of the fundamental biological components of eukaryotic life - indeed, every feature of the neochromosome may be defined down to the individual nucleobase. However, the flexibility incurred by designing such a large biological structure also presented a unique challenge due to its complexity. A series of design principles based on engineering concepts helped to rationalise this complexity and improve the overall design process. These included the presence of orthogonal DNA elements, a defined structural hierarchy, a control of replication and in-built modularity with the *rox* recombination system. The automation of tRNA flanking sequence assignment and sequence concatenation was also an efficient means of generating each tRNA array sequence.

The design principles were also intended to maintain functionality and maximise stability of the neochromosome. Orthogonal DNA elements, obtained from the genomes of yeast closely related to *S. cerevisiae,* were incorporated to reduce the risk of homologous recombination with the host genome. These genetic elements include origins of replication from *C. glabrata*, flanking sequences from *A. gossypii* and *E. cymbalariae* and

synthetic dual-block replication *Ter* sites from *Saccharomyces paradoxus*, *Saccharomyces uvarum*, *Saccharomyces mikatae*, *Saccharomyces pastorianus* and *Saccharomyces kudriavzevii*. It is likely that the tRNA neochromosome is the largest orthogonal DNA sequence yet designed and synthesised.

Further key to the process of design rationalisation was a defined structural hierarchy, (or 'hierarchy of abstraction'), intended to set distinct information boundaries between complex features of the neochromosome. This structural hierarchy consisted of small tRNA cassettes, larger tRNA arrays and a higher-order structure.

The first level of design hierarchy included small tRNA cassettes, replete with orthogonal tRNA flanking sequences designed to maintain a relative distance between tRNA genes, ensure transcriptional termination and reduce homology with the host genome. The presence of flanking *rox* recombination sites also presents a unique system to future study tRNA biology by systematically altering tDNA copy number. tRNA genes were assigned to flanking sequences based on an algorithm that ultimately formed the basis of Python programming scripts. Importantly, these scripts removed unwanted features from flanking sequences and concatenated these DNA elements to form tRNA arrays.

The resulting sixteen tRNA arrays provide a resource for Sc2.0 consortium members and were used to construct the neochromosome. The Chr2 and Chr12 tRNA arrays have already proven useful by reversing an up-regulation in the translational machinery in SynII (Shen *et al.*, 2017) and restoring a fitness defect in SynXII (Zhang *et al.*, 2017).

A truly unique feature of the neochromosome is a defined profile of DNA replication. By strategically placing origins of replication and termination sites, the higher-order structure of the neochromosome was designed to minimise replication stress and increase stability. Replication fork stall events are known to be polar in respect to tRNA

orientation relative to the replisome (Deshpande and Newlon, 1996), and so tRNA genes were orientated in a co-directional manner relative to the direction of the replication fork. The presence of replication termination cassettes was also designed to ensure that the replication fork terminates at defined regions, and prevent replication fork collision between tRNAs of opposing transcriptional orientation.

There are caveats recognised to the design process and a degree of risk is recognised. The optimal number and arrangement of origins of replication and replication termination sites couldn't be predicted before construction. In addition, convergent replication termination sites are yet to be fully tested in any eukaryotic system, and so a degree of risk was recognised by their presence. *Ter* sites are also known to be recombinogenic (Rothstein *et al.*, 2000), and may increase neochromosome instability. Finally, risk is observed in the systematic removal of tRNA introns, although the question of their essentiality is by no means certain.

However, the successful construction of the tRNA neochromosome demonstrated the feasibility of converting the *in silico* design into an *in vivo* object. Future validation of the above design decisions may be undertaken and ultimately altered, if necessary, for future neochromosome variants as part of the recursive Design-Build-Test cycle.

6.1.2. Comments on Construction Through the Inchworming Method

In vivo homologous recombination in yeast is an exceptionally powerful tool to assemble overlapping DNA fragments. The inchworming method applied homologous recombination to successfully construct the tRNA neochromosome from its constituent tRNA arrays.

The inchworming method should be placed in context with other large-scale DNA assembly methods. For example, the *Mycoplasma genitalium* genome was assembled using *in vivo* homologous recombination in yeast by combining a series of 25 overlapping DNA fragments in a single transformation (Gibson *et al.*, 2008). However, it is unclear if this approach could have been applied to construct the tRNA neochromosome - the unstable nature of tRNA genes and high proportion of false positives observed during construction likely render this strategy unfeasible. The inchworming method had the advantage of 'locking in' each DNA fragment through forced selection pressure by swapping the *URA3* and *LEU2* markers in a stepwise manner.

A further challenge observed during neochromosome construction was the redundancy of neochromosome DNA. Apart from the auxotrophic marker and the centromere, the wild-type cell has no requirement to maintain the presence of all 275 unstable tRNA cassettes. However, the stepwise manner of neochromosome construction, forced marker swapping and stringent screening process provided a reliable method to select for successful integration and the full neochromosome sequence. On some occasions, DNA fragments contained an origin of replication and selective marker: the presence of the *HIS3* marker on the neochromosome ensured selection against re-circularisation events.

Further advantages of the inchworming method included its flexibility: any new DNA fragment can be seamlessly introduced into the growing neochromosome, and therefore altered at will. Additionally, inchworming can be repeat *ad infinitum* until the carrying capacity of the cell for synthetic DNA is eventually reached. Finally, as previously-verified fragments are simply excised from their backbone using restriction digest, it was not necessary to fully sequence the neochromosome after each round of integration.

There are caveats for the inchworming method. The process was observed to be slow and required multiple rounds of screening with colony PCR after every round of transformation. During the latter rounds of construction, a significant proportion of isolates were observed to have lost neochromosome DNA, and so the efficiency of successful integration decreased to as low as 4%. Additionally, PCR-based methods are limited by their capacity to detect only the *presence* of DNA, and not its *conformation.* For example, the PCR tag approach did not detect the structural variations observed following deep sequencing in Chapter 5. Finally, the regular occurrence of an unwanted, additional selective marker necessitated its induced removal by re-streaking and replica plating after every round of inchworming.

However, it's unclear if any other DNA assembly method would have enabled the successful construction of the neochromosome. The inchworming method may be optimised and used to construct neochromosomes in future.

6.1.3. Discussion on Neochromosome Characterisation

The neochromosome is a unique biological structure that consists mostly of orthogonal genetic elements. Therefore, it was necessary to characterise these elements to ensure that they function as expected in *S. cerevisiae*. Characterisation efforts did not include work on the *rox* recombination system, which is not part of the scope of this doctoral study.

A series of orthogonal origins of replication and *Ter* sites were incorporated to define the neochromosome replication profile. Origins of replication from *C. glabrata* all functioned in *S. cerevisiae,* although the ARS *chrL-615* origin displayed a smaller colony size. Future studies on the nature of replication timing in the neochromosome may be performed: if the strength of this origin is observed to be too weak, it can simply be replaced with a

stronger version. Dual-block replication termination cassettes were also observed to function, although not with high efficiency. It should be noted that this assay was intended to be qualitative in nature and not quantitative – again, future studies on replication timing should also shed insight into their function in the neochromosome.

Systematically characterising all 275 synthetic tRNA genes, in addition to the native copy on the genome, presented a unique challenge. This challenge is compounded since synthetic and wild-type mature tRNAs can't be distinguished from each other. Only precursor tRNA molecules, based on the presence of short 5' leader and 3' trailer sequence, are capable of being mapped their locus on either the genome or neochromosome.

An alternate strategy was to characterise the function of synthetic tRNA genes on an individual level. Three, single-copy, essential tRNA genes were removed from the genome to study the effects of replacing their function with a synthetic variant. *SUP61, TRT2* and *TRR4* tRNA genes were integrated into the *HO* locus of strains lacking a wild-type copy of each. A yeast spotting assay was then performed - no observable fitness defect in terms of colony size was observed, demonstrating that the synthetic tRNA genes are fully functional. Finally, *SUP61* is the only single-copy, essential, tRNA gene containing an intron: it is interesting to note that the loss of the intron in the synthetic variant produced no observable effects on cellular fitness. This provides some evidence to suggest that tRNA introns may be non-essential.

Northern blot revealed the presence of an unexpected additional band for the synthetic SUP61 tRNA, likely representing an accumulation of the precursor transcript. To verify that this is indeed precursor accumulation, the transcriptional pol III inhibitor, thiolutin (Tipper, 1973), may be used in future to visualise a gradual processing of this band. If this

is an accumulation of the precursor transcript, a gradual loss over time should be observed in the presence of this inhibitor.

An experiment using a series of *SUP61* variants clearly show that the 5' flanking sequence from *A. gossypii* was responsible for the additional band. This result was very much unexpected: the endonuclease, RNase P, is responsible for processing the 5' leader sequence of the pre-tRNA molecule and recognises the T-loop of the tRNA itself, prior to cutting internally into the 5' leader (Kahle *et al.*, 1990). Previous work on sequence comparison shows that the 5' flanking sequence displays no common sequence element (Engelke *et al.*, 1985, Goodman *et al.*, 1977, Frank and Pace, 1998) and tRNAs with a 5' flanking sequence as little as a single nucleotide can be processed by RNase P (Kline *et al.*, 1981, Surratt *et al.*, 1990). Most the evidence in the literature strongly supports the hypothesis that the 5' leader sequence plays no role in substrate recognition by RNaseP.

However, results following Northern blot contrast with these previous findings. It is possible that the structure of the 5' and 3' leader and trailer sequences influence the processing efficiency of RNase P: previous work has shown that complementary pairing between the 5' and 3' sequences affects RNase P cleavage, and that these sequences may need to be separated for substrate recognition (Ziehler *et al.*, 2000).

An important question yet to be addressed is whether this potential processing defect is restricted to the synthetic SUP61 variant or present on all 275 tRNA genes of the neochromosome. In either case, colony morphology of strains housing synthetic variants of *SUP61*, *TRT2* and *TRR4* produced no observable fitness defect. In addition, preliminary systematic characterisation by external collaborators reveal that synthetic tRNAs appear to function normally (data not shown). While it remains possible that SUP61 is the only tRNA affected, further experimental work should be undertaken to verify that this is an

accumulation of the precursor form, elucidate the specific reason for its presence and ensure there are no systematic defects in tRNA processing.

A plate reader growth assay revealed a slow growth phenotype in strains housing the neochromosome. There are several potential sources of this growth defect, including delayed replication of the neochromosome or a metabolic burden caused by a doubling of the number of tRNA genes. Additionally, if the potential tRNA processing defects observed for SUP61 were present at a global level, negative fitness effects would likely accumulate as the neochromosome increased in size. However, preliminary (unpublished) results from global tRNA characterisation methods, described in further detail in Section 6.2.1, suggest that this is not be the case. Further work is required to systematically characterise tRNA expression and processing, and identify the reason for the slow growth phenotype.

6.1.4. PFGE and Sequencing of the tRNA Neochromosome Reveals Structural Variations

The neochromosome was constructed in two, independent, *S. cerevisiae* strain backgrounds: BY4741 and a strain housing a synthetic chromosome III, VI and right arm of chromosome IX. The early decision to construct the neochromosome these two strains may eventually prove invaluable: linearisation of the neochromosome constructed in the BY4741 strain background revealed a significant increase in neochromosome size. Subsequent preliminary sequencing results revealed a large, complex internal duplication and heterogeneity between isolates. This complex duplication may be unfeasible to repair; however, it might be possible to repair the comparatively minor structural variation of the neochromosome constructed in the triple-synthetic-chromosome background.

The remnant *ura3* marker in the *HO* locus of the triple-synthetic-chromosome strain background proved problematic for integration of the telomerator cassette. However, the chemical extraction and transfer of the neochromosome from the triple-synthetic-chromosome strain background to BY4741 not only circumvented this issue, but also demonstrated that the neochromosome is an intact, physical object. The chemical extraction method also demonstrates that neochromosomes of around 200 kb in size (and likely greater) can be efficiently transferred from one strain to another.

Following subsequent chemical extraction and transfer to BY4741, the neochromosome constructed in the triple-synthetic-chromosome strain background was again linearised at two regions distal to the centromere. Pulsed-field gel electrophoresis revealed that this version of the neochromosome produced a near-correct size following PFGE. However, sequencing shows that both linearised isolates contains structural variations, although not as significant as the massive duplication observed for the neochromosome constructed in BY4741. These structural variations include a duplication for the Chr13 tRNA array, the deletion of two high-copy tRNA cassettes and a large inversion.

These structural variations likely do not fundamentally affect neochromosome function or tRNA gene expression, although uneven replication generated by the inversion may alter replication timing: the *chrM-794* origin of replication is now required to replicate a ~20 kb and ~42 kb region on each side. To generate a near-perfect sequence, the neochromosome may be repaired in future.

6.1.5. Speculation on Potential Sources of Complex Neochromosome Structure Variations

The structural variations of the neochromosome in two independent strain backgrounds are unlikely to be coincidental. Potential sources of these structural variations are

discussed herein, and focus on the hypothesis that the circular structure of the neochromosome contributed to its instability. This hypothesis is based on the known instability of ring chromosomes in eukaryotes (discussed in Section 1.5.2. in Chapter 1) and evidence suggesting the structural variations occurred before linearisation. In **Figure 5.22**, independent isolates housing the neochromosome linearised at position B and position C both contain the same inversion and duplication. Additionally, *in vitro* linearisation of the neochromosome constructed in the BY4741 strain background (Section 5.2.2.2) revealed that the duplication event occurred prior to *in vivo* linearisation with the telomerator. Both results indicate that the structural variations occurred when the neochromosome existed in its circular form.

Although deep sequencing revealed that structural variations of the neochromosome occurred between *rox* recombination sites, it is proposed that the increased frequency of DNA breaks associated with tRNA gene transcription result in DNA repair through homologous recombination. The broken DNA ends will essentially seek homology for repair: due to the evenly-spaced and repetitive nature of *rox* recombination sites throughout the neochromosome, the probability of recombination between these sites will in turn increase.

However, DNA repair alone cannot account for the large duplications observed following PFGE and sequencing. The large-scale acquisition of DNA by the neochromosome can only be explained by the presence of an additional copy existing throughout the cell cycle, or, alternatively, by complementary pairing through sister chromatid exchange. Evidence supporting the former possibility includes the presence of a remnant selective marker observed during construction (**Figure 4.21**), suggesting the presence of an additional copy throughout the cell cycle. For the latter possibility, sister chromatid exchange between ring chromosome pairs are known to be a significant source of instability in eukaryotes (described in further detail in Chapter 1).

It's reasonable to speculate that the circular neochromosome displays similar unstable behaviour caused by either sister chromatid exchange or homologous recombination. Both scenarios are dependent on neochromosome copy number, although sister chromatid exchange may still occur in the presence of increased copy number. The potential events contributing to an acquisition of DNA in the neochromosome are summarised as follows and in **Figure 6.1**.

Sister Chromatid Exchange (Neochromosome Copy Number of 1 or Greater)

1. A single exchange of DNA between two neochromosome pairs may lead to the formation of a dicentric ring, resulting in an anaphase bridge and an uneven exchange of DNA during mitosis (**Figure 6.1B**).
2. A double exchange of DNA between two neochromosome pairs may result in catenation (interlocked ring) formation between the two circular neochromosome pairs, also resulting in an uneven transfer of DNA during mitosis (**Figure 6.1C**).

Homologous Recombination (Neochromosome Copy Number of 2 or Greater)

1. A single homologous recombination event between two neochromosome pairs may lead to the formation of a dicentric ring and associated instability (**Figure 6.1D**).
2. Homologous recombination between two neochromosome copies is independent of sister chromatid pairing and may occur at any location in the neochromosome body. A double homologous recombination event may result in the large-scale and uneven exchange of DNA, leading to the formation of two chromosomes of unequal size (**Figure 6.1E**). DNA catenation may also occur in this scenario.

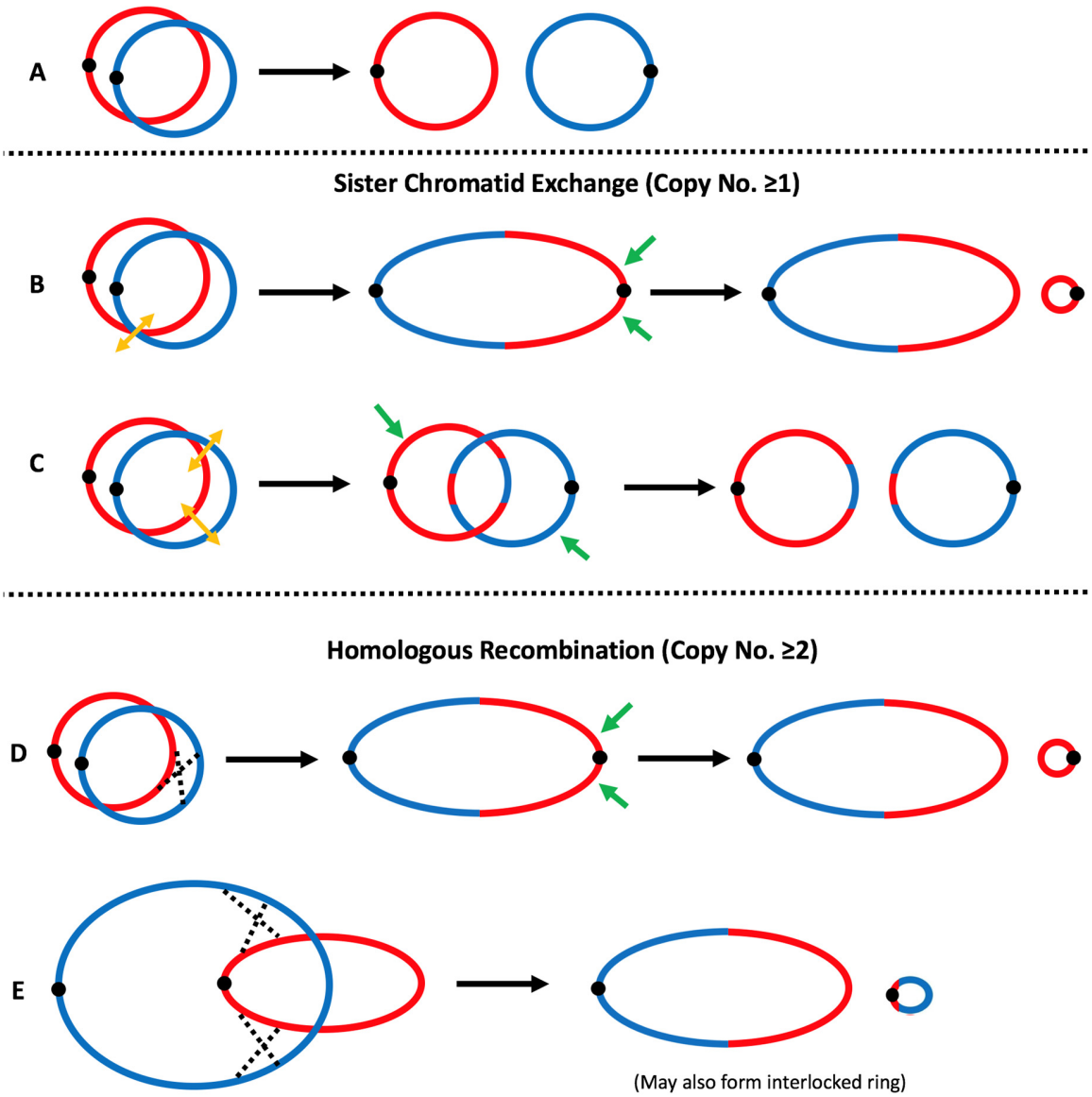


Figure 6.1: Potential models of circular neochromosome behaviour. The figure provides a graphical representation of the proposed model for neochromosome instability. In the above image, black dots denote the centromeres, yellow arrows indicate the sites of sister chromatid exchange, green arrows the sites of chromosome breakage and dotted lines the sites of homologous recombination. Copy No. refers to the number of neochromosome copies that exist within the cell prior to mitosis. To correlate with PFGE results, the above models are all circular based on the subsequent 'healing' of linear ends following DNA strand breaks.

[Figure legend continued on following page]

Figure 6.1 (Continued) (**A**) Normal separation will lead to two equal circular chromosomes during mitosis. (**B**) In the event of a single sister chromatid exchange, two circular chromosomes may fuse leading to the formation of a dicentric ring that is subsequently torn apart during cell division. An uneven breakpoint may also lead to the formation of both a large and small circular neochromosome in the respective daughter cells. (**C**) A dual sister chromatid exchange may lead to catenane (interlocked ring) formation of the two circular chromosome pairs, leading to uneven breakage and an imbalanced DNA products in the daughter cells. (**D**) A single homologous recombination event will behave similarly to a single sister chromatid exchange event, again leading to the formation of an unstable dicentric ring. (**E**) A dual homologous recombination event does not require sister chromatid alignment, and so any region of the neochromosome may recombine with its pair. The above image indicates a representative example of homologous recombination proximal to the centromere, producing a significant exchange of DNA and ultimately leading to the presence of one large and one small chromosome. A catenane may also form during homologous recombination, leading to DNA breakage and unequal segregation during mitosis. The above image was based on the model described by Yip (2015), and modified to account for the role of homologous recombination and neochromosome instability.

It should be noted that the exact mechanism of the increase in neochromosome size is unknown, and further work should be undertaken to identify the precise source of complex structural variations. It is unclear if the models above account for the *inversion* observed in **Figure 5.22,** although it should be noted that inversions, duplications and deletions have been observed following ring chromosome repair of telomere-proximal breaks in humans (Knijnenburg *et al.*, 2007, Rossi *et al.*, 2008). Additionally, PFGE results (**Figures 5.15** and **5.19**) indicate that the neochromosome still existed as a circular structure prior to linearisation with the telomerator, and so re-circularisation of the neochromosome following DNA breakage would be a necessary prerequisite to correlate with the above models.

To investigate chromosome fusion events, fluorescence *in situ* hybridisation (FISH) may be performed to directly visualise neochromosome behaviour through microscopy (Guilherme *et al.*, 2011, Knijnenburg *et al.*, 2007). To verify neochromosome copy

number, deep-sequencing, quantitative PCR or Multiplex Ligation-Dependent Probe Amplification (MLPA) may be performed (Knijnenburg *et al.*, 2007). Finally, if the circular structure of the neochromosome was responsible for the observed instability, subsequent linearisation with the telomerator may increase its stability.

6.2. Conclusion and Future Scope

The successful construction of the tRNA neochromosome in *S. cerevisiae* (albeit with structural variations) has not only demonstrated the feasibility of the approaches used, but the remarkable capacity of yeast to support an additional chromosome housing 275 unstable tRNA genes. This project also demonstrated the application of the recursive Design-Build-Test cycle and the development of new approaches to construct chromosomes *de novo*.

Overall, the tRNA neochromosome is a critical component of the Sc2.0 consortium - without a functional tRNA neochromosome, the synthetic cell will not be viable. The relocation of all 275 tRNA genes on a neochromosome should also, theoretically, result in a more stable synthetic genome structure and additionally present future opportunities to systematically study tRNA biology with the *rox* recombination system.

The construction of a tRNA neochromosome was always a high-risk and even radical undertaking. It's known that tRNA genes are hotspots for genomic instability, and so it couldn't have been predicted if their collation onto a neochromosome was even possible. However, this PhD project, at the very minimum, has demonstrated an ultimate proof-of-principle: an additional chromosome housing 275 tRNA genes is possible to construct.

The observed structural variations demonstrate the neochromosome may be unstable. If these regions are to be repaired, linearisation of the neochromosome using the

telomerator may increase stability. It's also important to note that the neochromosome contains largely redundant DNA: once the neochromosome is present in a synthetic yeast strain lacking all 275 wild-type tRNA genes, any issues of instability will be irrelevant so long as the neochromosome maintains viability.

The question of whether the neochromosome maintains a wild-type phenotype in a fully-synthetic strain is impossible to address at this stage. Although BY4741 was shown to tolerate synthetic copies of *SUP61*, *TRT2* and *TRR4*, global characterisation methods (including CRAC and tRNA sequencing, described further below) should provide a more systematic method of characterising tRNA genes of the neochromosome.

6.2.1. Collaborative Efforts Facilitate Global Neochromosome Characterisation and Potential for Future Work

The work undertaken in this doctoral study primarily focused on some initial aspects of neochromosome characterisation. Important collaboration efforts are intended to address aspects of characterisation in a more systematic manner. Additionally, follow-up experimental work will be performed to address biological questions of chromosome behaviour, such as segregation and replication. Finally, further work will be undertaken to characterise the Dre-*rox* recombination system.

Although Northern blot is highly sensitive measure of characterising tRNA expression, it is low-throughput. To systematically study tRNA genes of the neochromosome, CRAC (UV cross-linking and analysis of cDNA) and tRNA sequencing will be performed. These methods allow the mapping of nascent pre-tRNAs based on the transient trailer and leader sequences, allowing the differentiation of neochromosome tRNAs from the genome. Preliminary data from these two technologies reveal that neochromosome

tRNA expression levels are approximately 87% comparable to wild-type, and no apparent systematic defects are observed (data not shown).

CRAC may be used to map nascent tRNA transcripts transcribed by RNA polymerase III, and provide a global measure of tRNA transcription (Turowski *et al.*, 2016). This work is currently being undertaken by the Tollervey group (Edinburgh, UK).

A bespoke method of tRNA sequencing is currently being developed by the Steinmetz group (Heildelberg, Germany). tRNA molecules, by their very nature, are not amenable to traditional RNA sequencing methods due to their highly-folded structure and hard-stop modifications that block reverse transcriptase (Wilusz, 2015). However, two recently described methods have been used to remove tRNA methylation and facilitate reverse transcription: ARM-seq (Cozen *et al.*, 2015) uses an adaptor ligation-based method and DM-tRNA-seq (Zheng *et al.*, 2015) utilises a highly processive group II intron reverse transcriptase, although the latter is yet to be tested in yeast.

Collaboration efforts also include work performed by the Antequera group (Salamanca, Spain) on the mapping of nucleosomes throughout the neochromosome (Gonzalez *et al.*, 2016), RNA sequencing by the Beijing Genomics Institute (Shenzhen, China) and work by the Nieduszynski group (Oxford, UK) to characterise DNA replication with deep-sequencing (Muller *et al.*, 2014, Hawkins *et al.*, 2013). The latter is important to characterise replication origins and *Ter* sites of the neochromosome.

tRNA genes may contribute to the general organisation of the genome in the nuclear space. Research suggests that tRNA genes cluster in the nucleolus (Thompson *et al.*, 2003), although some research suggests that tRNA genes cluster around the nuclear pores during M-phase (Chen and Gartenberg, 2014) or, in some instances, around the centromeres (Duan *et al.*, 2010). Chromosome-conformation-capture methods (Hi-C)

(Mercy *et al.*, 2017) or, alternatively, fluorescence in situ hybridisation (FISH) will be used to address the question of intracellular locality. HI-C analysis of the neochromosome will be performed by the Koszul group (Paris, France).

Further follow-up experimental work includes segregation studies that will be performed using the TetR-TetO system (Michaelis *et al.*, 1997). TetO repeats can be integrated near the centromere, and once tagged with GFP, microscopic analysis will provide an indication of neochromosome segregation behaviour during mitosis. Finally, further experimental work will be performed to characterise the efficiency of the Dre *rox* recombination system in the context of the neochromosome. This will be undertaken by Ms. Wei Liu (Edinburgh, UK).

6.2.2. Neochromosome Version 2.0?

Due to its importance as part of the Sc2.0 consortium, the neochromosome may be revisited and improved in future as part of the Design-Build-Test cycle. With rapidly increasing capabilities in DNA synthesis technology and potential improvements to the inchworming method, a new version or versions of the tRNA neochromosome (neochromosome version 2.0) may be constructed based on lessons learned in this study. For example, characterising DNA replication with deep sequencing will provide an indication of the optimal number of origins of replication and replication termination sites.

In all cases observed, structural variations of the neochromosome (including deletions and structural variations) were caused by recombination between *rox* recombination sites. It is possible that the removal of *rox* sites will facilitate a more stable version of the neochromosome. Finally, it is unclear at this stage if the presence of synthetic telomeres produce a silencing effect on tRNA gene expression. If this is the case, orthogonal sub-

telomeric anti-silencing regions (STARs) (Fourel *et al.*, 1999) may be used to reduce the effects of telomere-mediated tRNA gene silencing.

6.3. Concluding Statement

The tRNA neochromosome is an important, if not essential, component of the international Sc2.0 consortium. It is hoped the findings in this doctoral study will enhance the aims of the Sc2.0 consortium through current or future iterations of the neochromosome and enable the systematic study of tRNA genetics with the *rox* recombination system. At the time of writing, this project is likely the first time a true designer eukaryotic neochromosome has been designed and constructed *de novo.* This project has also presented a unique opportunity to design one of the fundamental building blocks of eukaryotic life in detail down to the individual nucleobase, demonstrating the great power and potential to engineer synthetic chromosomes *de novo*.

References

ABBAS, A. C. & NIELSEN, J. 2016. Synthetic yeast as the new frontier in evolutionary developments in biology. *FEMS Yeast Res,* 8, 16.

ADMIRE, A., SHANKS, L., DANZL, N., WANG, M., WEIER, U., STEVENS, W., HUNT, E. & WEINERT, T. 2006. Cycles of chromosome instability are associated with a fragile site and are increased by defects in DNA replication and checkpoint controls in yeast. *Genes Dev,* 20**,** 159-73.

AGARWAL, K. L., BUCHI, H., CARUTHERS, M. H., GUPTA, N., KHORANA, H. G., KLEPPE, K., KUMAR, A., OHTSUKA, E., RAJBHANDARY, U. L., VAN DE SANDE, J. H., SGARAMELLA, V., WEBER, H. & YAMADA, T. 1970. Total Synthesis of the Gene for an Alanine Transfer Ribonucleic Acid from Yeast. *Nature,* 227**,** 27-34.

AGUILERA, A. 2002. The connection between transcription and genomic instability. *Embo j,* 21**,** 195-201.

AGUILERA, A. & GARCIA-MUSE, T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell,* 46**,** 115-24.

AGUILERA, A. & GARCIA-MUSE, T. 2013. Causes of genome instability. *Annu Rev Genet,* 47**,** 1-32.

ALLISON, D. S. & HALL, B. D. 1985. Effects of alterations in the 3' flanking sequence on *in vivo* and *in vitro* expression of the yeast *SUP4-o* tRNATyr gene. *EMBO J,* 4**,** 2657-64.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

AN, W. & CHIN, J. W. 2009. Synthesis of orthogonal transcription-translation networks. *Proc Natl Acad Sci U S A,* 106**,** 8477-82.

ANASTASSIADIS, K., FU, J., PATSCH, C., HU, S., WEIDLICH, S., DUERSCHKE, K., BUCHHOLZ, F., EDENHOFER, F. & STEWART, A. F. 2009. Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli,* mammalian cells and mice. *Dis Model Mech,* 2**,** 508-15.

ANNALURU, N., MULLER, H., MITCHELL, L. A., RAMALINGAM, S., STRACQUADANIO, G., RICHARDSON, S. M., DYMOND, J. S., KUANG, Z., SCHEIFELE, L. Z., COOPER, E. M., CAI, Y., ZELLER, K., AGMON, N., HAN, J. S., HADJITHOMAS, M., TULLMAN, J., CARAVELLI, K., CIRELLI, K., GUO, Z., LONDON, V., YELURU, A., MURUGAN, S., KANDAVELOU, K., AGIER, N., FISCHER, G., YANG, K., MARTIN, J. A., BILGEL, M., BOHUTSKI, P., BOULIER, K. M., CAPALDO, B. J., CHANG, J., CHAROEN, K., CHOI, W. J., DENG, P., DICARLO, J. E., DOONG, J., DUNN, J., FEINBERG, J. I., FERNANDEZ, C., FLORIA, C. E., GLADOWSKI, D., HADIDI, P., ISHIZUKA, I., JABBARI, J., LAU, C. Y., LEE, P. A., LI, S., LIN, D., LINDER, M. E., LING, J., LIU, J., LIU, J., LONDON, M., MA, H., MAO, J., MCDADE, J. E., MCMILLAN, A., MOORE, A. M., OH, W. C., OUYANG, Y., PATEL, R., PAUL, M., PAULSEN, L. C., QIU, J., RHEE, A., RUBASHKIN, M. G., SOH, I. Y., SOTUYO, N. E., SRINIVAS, V., SUAREZ, A., WONG, A., WONG, R., XIE, W. R., XU, Y., YU, A. T., KOSZUL, R., BADER, J. S., BOEKE, J. D. &

CHANDRASEGARAN, S. 2014. Total synthesis of a functional designer eukaryotic chromosome. *Science,* 344**,** 55-8.
APARICIO, O. M., WEINSTEIN, D. M. & BELL, S. P. 1997. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell,* 91**,** 59-69.
BARRE, F. X., SOBALLE, B., MICHEL, B., AROYO, M., ROBERTSON, M. & SHERRATT, D. 2001. Circles: the replication-recombination-chromosome segregation connection. *Proc Natl Acad Sci U S A,* 98**,** 8189-95.
BERKOVITZ, G., STAMBERG, J., PLOTNICK, L. P. & LANES, R. 1983. Turner syndrome patients with a ring X chromosome. *Clin Genet,* 23**,** 447-53.
BERTRAND, E., HOUSER-SCOTT, F., KENDALL, A., SINGER, R. H. & ENGELKE, D. R. 1998. Nucleolar localization of early tRNA processing. *Genes Dev,* 12**,** 2463-8.
BOEKE, J. D., CHURCH, G., HESSEL, A., KELLEY, N. J., ARKIN, A., CAI, Y., CARLSON, R., CHAKRAVARTI, A., CORNISH, V. W., HOLT, L., ISAACS, F. J., KUIKEN, T., LAJOIE, M., LESSOR, T., LUNSHOF, J., MAURANO, M. T., MITCHELL, L. A., RINE, J., ROSSER, S., SANJANA, N. E., SILVER, P. A., VALLE, D., WANG, H., WAY, J. C. & YANG, L. 2016. The Genome Project-Write. *Science,* 353**,** 126-7.
BRAGLIA, P., PERCUDANI, R. & DIECI, G. 2005. Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III. *J Biol Chem,* 280**,** 19551-62.
BURGERS, P. M. J. & PERCIVAL, K. J. 1987. Transformation of yeast spheroplasts without cell fusion. *Analytical Biochemistry,* 163**,** 391-397.
CHAKSHUSMATHI, G., KIM, S. D., RUBINSON, D. A. & WOLIN, S. L. 2003. A La protein requirement for efficient pre-tRNA folding. *EMBO J,* 22**,** 6562-72.
CHAMBERLAIN, J. R., LEE, Y., LANE, W. S. & ENGELKE, D. R. 1998. Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes Dev,* 12**,** 1678-90.
CHAN, Y. A., HIETER, P. & STIRLING, P. C. 2014. Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet,* 30**,** 245-53.
CHANG, E. Y. & STIRLING, P. C. 2017. Replication Fork Protection Factors Controlling R-Loop Bypass and Suppression. *Genes,* 8.
CHARBIN, A., BOUCHOUX, C. & UHLMANN, F. 2014. Condensin aids sister chromatid decatenation by topoisomerase II. *Nucleic Acids Res,* 42**,** 340-8.
CHEN, J. Y. & MARTIN, N. C. 1988. Biosynthesis of tRNA in yeast mitochondria. An endonuclease is responsible for the 3'-processing of tRNA precursors. *J Biol Chem,* 263**,** 13677-82.
CHEN, M. & GARTENBERG, M. R. 2014. Coordination of tRNA transcription with export at nuclear pore complexes in budding yeast. *Genes & Development,* 28**,** 959-970.
CHEN, Y., BECK, A., DAVENPORT, C., CHEN, Y., SHATTUCK, D. & TAVTIGIAN, S. V. 2005. Characterization of *TRZ1*, a yeast homolog of the human candidate prostate cancer susceptibility gene *ELAC2* encoding tRNase Z. *BMC Mol Biol,* 6**,** 12.

CLARK, M. W. & ABELSON, J. 1987. The subnuclear localization of tRNA ligase in yeast. *J Cell Biol,* 105**,** 1515-26.
CLARKE, L. 1990. Centromeres of budding and fission yeasts. *Trends Genet,* 6**,** 150-4.
COHEN, S. N., CHANG, A. C., BOYER, H. W. & HELLING, R. B. 1973. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A,* 70**,** 3240-4.
COOPER, E. M., MULLER, H., CHANDRASEGARAN, S., BADER, J. S. & BOEKE, J. D. 2012. The Build-a-Genome course. *Methods Mol Biol,* 852**,** 273-83.
COPELA, L. A., CHAKSHUSMATHI, G., SHERRER, R. L. & WOLIN, S. L. 2006. The La protein functions redundantly with tRNA modification enzymes to ensure tRNA structural stability. *RNA,* 12**,** 644-54.
COPELA, L. A., FERNANDEZ, C. F., SHERRER, R. L. & WOLIN, S. L. 2008. Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. *RNA,* 14**,** 1214-1227.
COZEN, A. E., QUARTLEY, E., HOLMES, A. D., HRABETA-ROBINSON, E., PHIZICKY, E. M. & LOWE, T. M. 2015. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods,* 12**,** 879-84.
CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227**,** 561-3.
CULVER, G. M., MCCRAITH, S. M., CONSAUL, S. A., STANFORD, D. R. & PHIZICKY, E. M. 1997. A 2'-phosphotransferase implicated in tRNA splicing is essential in *Saccharomyces cerevisiae*. *J Biol Chem,* 272**,** 13203-10.
DANI, G. M. & ZAKIAN, V. A. 1983. Mitotic and meiotic stability of linear plasmids in yeast. *Proc Natl Acad Sci U S A,* 80**,** 3406-10.
DE LANGE, T. 2009. How telomeres solve the end-protection problem. *Science,* 326**,** 948-52.
DESHPANDE, A. M. & NEWLON, C. S. 1992. The ARS consensus sequence is required for chromosomal origin function in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology,* 12**,** 4305-4313.
DESHPANDE, A. M. & NEWLON, C. S. 1996. DNA replication fork pause sites dependent on transcription. *Science,* 272**,** 1030-3.
DEVENISH, R. J. & NEWLON, C. S. 1982. Isolation and characterization of yeast ring chromosome III by a method applicable to other circular DNAs. *Gene,* 18**,** 277-88.
DHUNGEL, N. & HOPPER, A. K. 2012. Beyond tRNA cleavage: novel essential function for yeast tRNA splicing endonuclease unrelated to tRNA processing. *Genes Dev,* 26**,** 503-14.
DIETRICH, F. S., VOEGELI, S., BRACHAT, S., LERCH, A., GATES, K., STEINER, S., MOHR, C., POHLMANN, R., LUEDI, P., CHOI, S., WING, R. A., FLAVIER, A., GAFFNEY, T. D. & PHILIPPSEN, P. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science,* 304**,** 304-7.

DIETRICH, F. S., VOEGELI, S., KUO, S. & PHILIPPSEN, P. 2013. Genomes of Ashbya fungi isolated from insects reveal four mating-type loci, numerous translocations, lack of transposons, and distinct gene duplications. *G3 (Bethesda),* 3**,** 1225-39.

DORMAN, C. J. & DORMAN, M. J. 2016. DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophysical Reviews,* 8**,** 209-220.

DUAN, Z., ANDRONESCU, M., SCHUTZ, K., MCILWAIN, S., KIM, Y. J., LEE, C., SHENDURE, J., FIELDS, S., BLAU, C. A. & NOBLE, W. S. 2010. A three-dimensional model of the yeast genome. *Nature,* 465**,** 363-7.

DYMOND, J. S., RICHARDSON, S. M., COOMBES, C. E., BABATZ, T., MULLER, H., ANNALURU, N., BLAKE, W. J., SCHWERZMANN, J. W., DAI, J., LINDSTROM, D. L., BOEKE, A. C., GOTTSCHLING, D. E., CHANDRASEGARAN, S., BADER, J. S. & BOEKE, J. D. 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature,* 477**,** 471-6.

EL HAGE, A., FRENCH, S. L., BEYER, A. L. & TOLLERVEY, D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev,* 24**,** 1546-58.

EL HAGE, A., WEBB, S., KERR, A. & TOLLERVEY, D. 2014. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet,* 10**,** e1004716.

ENGELKE, D. R., GEGENHEIMER, P. & ABELSON, J. 1985. Nucleolytic processing of a tRNAArg-tRNAAsp dimeric precursor by a homologous component from *Saccharomyces cerevisiae*. *J Biol Chem,* 260**,** 1271-9.

ENGLER, C., KANDZIA, R. & MARILLONNET, S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One,* 3**,** e3647.

FITZGERALD-HAYES, M., CLARKE, L. & CARBON, J. 1982. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell,* 29**,** 235-44.

FOUREL, G., REVARDEL, E., KOERING, C. E. & GILSON, E. 1999. Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *EMBO J,* 18**,** 2522-37.

FRAGKOS, M. & NAIM, V. 2017. Rescue from replication stress during mitosis. *Cell Cycle,* 16**,** 613-633.

FRANK, D. N. & PACE, N. R. 1998. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem,* 67**,** 153-80.

GAILLARD, H. & AGUILERA, A. 2016. Transcription as a Threat to Genome Integrity. *Annu Rev Biochem,* 85**,** 291-317.

GARNEAU, J. E., DUPUIS, M. E., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADAN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468**,** 67-71.

GEBHART, E. 2008. Ring chromosomes in human neoplasias. *Cytogenet Genome Res,* 121**,** 149-73.

GIBSON, D. G., BENDERS, G. A., AXELROD, K. C., ZAVERI, J., ALGIRE, M. A., MOODIE, M., MONTAGUE, M. G., VENTER, J. C., SMITH, H. O. & HUTCHISON, C. A., 3RD 2008. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A,* 105**,** 20404-9.

GIBSON, D. G., GLASS, J. I., LARTIGUE, C., NOSKOV, V. N., CHUANG, R. Y., ALGIRE, M. A., BENDERS, G. A., MONTAGUE, M. G., MA, L., MOODIE, M. M., MERRYMAN, C., VASHEE, S., KRISHNAKUMAR, R., ASSAD-GARCIA, N., ANDREWS-PFANNKOCH, C., DENISOVA, E. A., YOUNG, L., QI, Z. Q., SEGALL-SHAPIRO, T. H., CALVEY, C. H., PARMAR, P. P., HUTCHISON, C. A., 3RD, SMITH, H. O. & VENTER, J. C. 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science,* 329**,** 52-6.

GIBSON, D. G., YOUNG, L., CHUANG, R.-Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth,* 6**,** 343-345.

GIESE, B., KOENIGSTEIN, S., WIGGER, H., SCHMIDT, J. C. & VON GLEICH, A. 2013. Rational Engineering Principles in Synthetic Biology: A Framework for Quantitative Analysis and an Initial Assessment. *Biological Theory,* 8**,** 324-333.

GIETZ, R. D. & WOODS, R. A. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol,* 350**,** 87-96.

GISSELSSON, D. 2002. Ring chromosomes: vicious circles at the end, beginning of life. *Atlas Genet Cytogenet Oncol Haematol.,* 6**,** 62-69.

GISSELSSON, D., PETTERSSON, L., HOGLUND, M., HEIDENBLAD, M., GORUNOVA, L., WIEGANT, J., MERTENS, F., DAL CIN, P., MITELMAN, F. & MANDAHL, N. 2000. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc Natl Acad Sci U S A,* 97**,** 5357-62.

GOFFEAU, A., BARRELL, B. G., BUSSEY, H., DAVIS, R. W., DUJON, B., FELDMANN, H., GALIBERT, F., HOHEISEL, J. D., JACQ, C., JOHNSTON, M., LOUIS, E. J., MEWES, H. W., MURAKAMI, Y., PHILIPPSEN, P., TETTELIN, H. & OLIVER, S. G. 1996. Life with 6000 genes. *Science,* 274**,** 546-567.

GOMEZ-GONZALEZ, B., GARCIA-RUBIO, M., BERMEJO, R., GAILLARD, H., SHIRAHIGE, K., MARIN, A., FOIANI, M. & AGUILERA, A. 2011. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *Embo j,* 30**,** 3106-19.

GONZALEZ, S., GARCIA, A., VAZQUEZ, E., SERRANO, R., SANCHEZ, M., QUINTALES, L. & ANTEQUERA, F. 2016. Nucleosomal signatures impose nucleosome positioning in coding and noncoding sequences in the genome. *Genome Res,* 26**,** 1532-1543.

GOODMAN, H. M., OLSON, M. V. & HALL, B. D. 1977. Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor SUP4-o. *Proc Natl Acad Sci U S A,* 74**,** 5453-7.

GOTTIPATI, P., CASSEL, T. N., SAVOLAINEN, L. & HELLEDAY, T. 2008. Transcription-Associated Recombination Is Dependent on Replication in Mammalian Cells. *Molecular and Cellular Biology,* 28**,** 154-164.

GREEN, R. & ROGERS, E. J. 2013. Chemical Transformation of *E. coli*. *Methods in enzymology,* 529**,** 329-336.

GUILHERME, R. S., MELONI, V. F., KIM, C. A., PELLEGRINO, R., TAKENO, S. S., SPINNER, N. B., CONLIN, L. K., CHRISTOFOLINI, D. M., KULIKOWSKI, L. D. & MELARAGNO, M. I. 2011. Mechanisms of ring chromosome formation, ring instability and clinical consequences. *BMC Med Genet,* 12**,** 171.

HAGE, A. E. & HOUSELEY, J. 2013. Resolution of budding yeast chromosomes using pulsed-field gel electrophoresis. *Methods Mol Biol,* 1054**,** 195-207.

HANI, J. & FELDMANN, H. 1998. tRNA genes and retroelements in the yeast genome. *Nucleic Acids Res,* 26**,** 689-96.

HARTMAN, H. & SMITH, T. F. 2014. The evolution of the ribosome and the genetic code. *Life,* 4**,** 227-49.

HAWKINS, M., RETKUTE, R., MULLER, C. A., SANER, N., TANAKA, T. U., DE MOURA, A. P. & NIEDUSZYNSKI, C. A. 2013. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep,* 5**,** 1132-41.

HEALE, S. M., STATEVA, L. I. & OLIVER, S. G. 1994. Introduction of YACs into intact yeast cells by a procedure which shows low levels of recombinagenicity and co-transformation. *Nucleic Acids Research,* 22**,** 5011-5015.

HEINEMANN, M. & PANKE, S. 2006. Synthetic biology--putting engineering into biology. *Bioinformatics,* 22**,** 2790-9.

HELMRICH, A., BALLARINO, M., NUDLER, E. & TORA, L. 2013. Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol,* 20**,** 412-8.

HIELM, S., BJORKROTH, J., HYYTIA, E. & KORKEALA, H. 1998. Genomic analysis of *Clostridium botulinum* group II by pulsed-field gel electrophoresis. *Appl Environ Microbiol,* 64**,** 703-8.

HIETER, P., MANN, C., SNYDER, M. & DAVIS, R. W. 1985. Mitotic stability of yeast chromosomes: A colony color assay that measures nondisjunction and chromosome loss. *Cell,* 40**,** 381-392.

HOPPER, A. K. 2013. Transfer RNA Post-Transcriptional Processing, Turnover, and Subcellular Dynamics in the Yeast Saccharomyces cerevisiae. *Genetics,* 194**,** 43-67.

HOPPER, A. K., PAI, D. A. & ENGELKE, D. R. 2010. Cellular dynamics of tRNAs and their genes. *FEBS Lett,* 584**,** 310-7.

HOPPER, A. K. & PHIZICKY, E. M. 2003. tRNA transfers to the limelight. *Genes Dev,* 17**,** 162-80.

HUANG, B., LU, J. & BYSTROM, A. S. 2008. A genome-wide screen identifies genes required for formation of the wobble nucleoside 5-methoxycarbonylmethyl-2-thiouridine in *Saccharomyces cerevisiae*. *RNA,* 14**,** 2183-94.

HUANG, Y. & MARAIA, R. J. 2001. Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human. *Nucleic Acids Res,* 29**,** 2675-90.

ITAYA, M., TSUGE, K., KOIZUMI, M. & FUJITA, K. 2005. Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc Natl Acad Sci U S A,* 102**,** 15971-6.

IVANCIC-BACE, I., HOWARD, J. A. & BOLT, E. L. 2012. Tuning in to interference: R-loops and cascade complexes in CRISPR immunity. *J Mol Biol,* 422**,** 607-16.

IVESSA, A. S., LENZMEIER, B. A., BESSLER, J. B., GOUDSOUZIAN, L. K., SCHNAKENBERG, S. L. & ZAKIAN, V. A. 2003. The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol Cell,* 12**,** 1525-36.

JI, H., MOORE, D. P., BLOMBERG, M. A., BRAITERMAN, L. T., VOYTAS, D. F., NATSOULIS, G. & BOEKE, J. D. 1993. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell,* 73**,** 1007-18.

JOHNSON, P. F. & ABELSON, J. 1983. The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. *Nature,* 302**,** 681-687.

KAHLE, D., WEHMEYER, U. & KRUPP, G. 1990. Substrate recognition by RNase P and by the catalytic M1 RNA: identification of possible contact points in pre-tRNAs. *EMBO J,* 9**,** 1929-37.

KANAI, A. 2013. Molecular Evolution of Disrupted Transfer RNA Genes and Their Introns in Archaea. *In:* PONTAROTTI, P. (ed.) *Evolutionary Biology: Exobiology and Evolutionary Mechanisms.* Berlin, Heidelberg: Springer Berlin Heidelberg.

KARKUSIEWICZ, I., TUROWSKI, T. W., GRACZYK, D., TOWPIK, J., DHUNGEL, N., HOPPER, A. K. & BOGUTA, M. 2011. Maf1 protein, repressor of RNA polymerase III, indirectly affects tRNA processing. *J Biol Chem,* 286**,** 39478-88.

KAWAI, S., HASHIMOTO, W. & MURATA, K. 2010. Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism. *Bioeng Bugs,* 1**,** 395-403.

KIM, J. M., VANGURI, S., BOEKE, J. D., GABRIEL, A. & VOYTAS, D. F. 1998. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Res,* 8**,** 464-78.

KLINE, L., NISHIKAWA, S. & SOLL, D. 1981. Partial purification of RNase P from *Schizosaccharomyces pombe*. *J Biol Chem,* 256**,** 5058-63.

KNIJNENBURG, J., VAN HAERINGEN, A., HANSSON, K. B., LANKESTER, A., SMIT, M. J., BELFROID, R. D., BAKKER, E., ROSENBERG, C., TANKE, H. J. & SZUHAI, K. 2007. Ring chromosome formation as a novel escape mechanism in patients with inverted duplication and terminal deletion. *Eur J Hum Genet,* 15**,** 548-55.

KÖNIG, H., FRANK, D., HEIL, R. & COENEN, C. 2013. Synthetic Genomics and Synthetic Biology Applications Between Hopes and Concerns. *Current Genomics,* 14**,** 11-24.
KUPIEC, M. 2014. Biology of telomeres: lessons from budding yeast. *FEMS Microbiol Rev,* 38**,** 144-71.
KUTTER, C., BROWN, G. D., GONCALVES, A., WILSON, M. D., WATT, S., BRAZMA, A., WHITE, R. J. & ODOM, D. T. 2011. Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nat Genet,* 43**,** 948-955.
LEMOINE, F. J., DEGTYAREVA, N. P., LOBACHEV, K. & PETES, T. D. 2005. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell,* 120**,** 587-98.
LERAT, E. & CAPY, P. 1999. Retrotransposons and retroviruses: analysis of the envelope gene. *Mol Biol Evol,* 16**,** 1198-207.
LIN, Y., DENT, S. Y., WILSON, J. H., WELLS, R. D. & NAPIERALA, M. 2010. R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A,* 107**,** 692-7.
LINDSTROM, D. L. & GOTTSCHLING, D. E. 2009. The mother enrichment program: a genetic system for facile replicative life span analysis in *Saccharomyces cerevisiae*. *Genetics,* 183**,** 413-22, 1SI-13SI.
LIU, R., BASSALO, M. C., ZEITOUN, R. I. & GILL, R. T. 2015. Genome scale engineering techniques for metabolic engineering. *Metab Eng,* 32**,** 143-54.
MADIREDDY, A., KOSIYATRAKUL, SETTAPONG T., BOISVERT, REBECCA A., HERRERA-MOYANO, E., GARCÍA-RUBIO, MARÍA L., GERHARDT, J., VUONO, ELIZABETH A., OWEN, N., YAN, Z., OLSON, S., AGUILERA, A., HOWLETT, NIALL G. & SCHILDKRAUT, CARL L. 2016. FANCD2 Facilitates Replication through Common Fragile Sites. *Molecular Cell,* 64**,** 388-404.
MANDEL, M. & HIGA, A. 1970. Calcium-dependent bacteriophage DNA infection. *J Mol Biol,* 53**,** 159-62.
MARAIA, R. J. & LAMICHHANE, T. N. 2011. 3' processing of eukaryotic precursor tRNAs. *Wiley Interdiscip Rev RNA,* 2**,** 362-75.
MCCLINTOCK, B. 1938. The Production of Homozygous Deficient Tissues with Mutant Characteristics by Means of the Aberrant Mitotic Behavior of Ring-Shaped Chromosomes. *Genetics,* 23**,** 315-76.
MCDERMOTT, A., VOYCE, M. A. & ROMAIN, D. 1977. Ring chromosome 4. *J Med Genet,* 14**,** 228-32.
MERCY, G., MOZZICONACCI, J., SCOLARI, V. F., YANG, K., ZHAO, G., THIERRY, A., LUO, Y., MITCHELL, L. A., SHEN, M., SHEN, Y., WALKER, R., ZHANG, W., WU, Y., XIE, Z.-X., LUO, Z., CAI, Y., DAI, J., YANG, H., YUAN, Y.-J., BOEKE, J. D., BADER, J. S., MULLER, H. & KOSZUL, R. 2017. 3D organization of synthetic and scrambled chromosomes. *Science,* 355**,** 1050.
MICHAELIS, C., CIOSK, R. & NASMYTH, K. 1997. Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids. *Cell,* 91**,** 35-45.

MIECZKOWSKI, P. A., LEMOINE, F. J. & PETES, T. D. 2006. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. *DNA Repair,* 5**,** 1010-1020.

MITCHELL, L. A. & BOEKE, J. D. 2014. Circular permutation of a synthetic eukaryotic chromosome with the telomerator. *Proc Natl Acad Sci U S A,* 111**,** 17003-10.

MITCHELL, L. A., WANG, A., STRACQUADANIO, G., KUANG, Z., WANG, X., YANG, K., RICHARDSON, S., MARTIN, J. A., ZHAO, Y., WALKER, R., LUO, Y., DAI, H., DONG, K., TANG, Z., YANG, Y., CAI, Y., HEGUY, A., UEBERHEIDE, B., FENYÖ, D., DAI, J., BADER, J. S. & BOEKE, J. D. 2017. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science,* 355**,** 1045.

MORI, S., KAJITA, T., ENDO, T. & YOSHIHISA, T. 2011. The intron of tRNA-Trp(CCA) is dispensable for growth and translation of *Saccharomyces cerevisiae*. *RNA,* 17**,** 1760-1769.

MORTIMER, R. K., GAME, J. C., BELL, M. & REBECCA CONTOPOULOU, C. 1990. Use of pulsed-field gel electrophoresis to study the chromosomes of *Saccharomyces* and other yeasts. *Methods,* 1**,** 169-179.

MULLER, C. A., HAWKINS, M., RETKUTE, R., MALLA, S., WILSON, R., BLYTHE, M. J., NAKATO, R., KOMATA, M., SHIRAHIGE, K., DE MOURA, A. P. & NIEDUSZYNSKI, C. A. 2014. The dynamics of genome replication using deep sequencing. *Nucleic Acids Res,* 42**,** e3.

MURRAY, A. W. & SZOSTAK, J. W. 1983. Pedigree analysis of plasmid segregation in yeast. *Cell,* 34**,** 961-970.

MURRAY, A. W. & SZOSTAK, J. W. 1985. Chromosome segregation in mitosis and meiosis. *Annu Rev Cell Biol,* 1**,** 289-315.

NGUYEN, V. C., CLELLAND, B. W., HOCKMAN, D. J., KUJAT-CHOY, S. L., MEWHORT, H. E. & SCHULTZ, M. C. 2010. Replication stress checkpoint signaling controls tRNA gene transcription. *Nat Struct Mol Biol,* 17**,** 976-81.

NOSKOV, V. N., CHUANG, R. Y., GIBSON, D. G., LEEM, S. H., LARIONOV, V. & KOUPRINA, N. 2011. Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat Protoc,* 6**,** 89-96.

OSBOURN, A. E., O'MAILLE, P. E., ROSSER, S. J. & LINDSEY, K. 2012. Synthetic biology. *New Phytologist,* 196**,** 671-677.

PERCUDANI, R., PAVESI, A. & OTTONELLO, S. 1997. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J Mol Biol,* 268**,** 322-30.

PETTERSSON, E., LUNDEBERG, J. & AHMADIAN, A. 2009. Generations of sequencing technologies. *Genomics,* 93**,** 105-11.

PHIZICKY, E. M., SCHWARTZ, R. C. & ABELSON, J. 1986. *Saccharomyces cerevisiae* tRNA ligase. Purification of the protein and isolation of the structural gene. *J Biol Chem,* 261**,** 2978-86.

PLUTA, K., LEFEBVRE, O., MARTIN, N. C., SMAGOWICZ, W. J., STANFORD, D. R., ELLIS, S. R., HOPPER, A. K., SENTENAC, A. & BOGUTA, M. 2001. Maf1p, a negative effector of RNA polymerase III in *Saccharomyces cerevisiae*. *Mol Cell Biol,* 21**,** 5031-40.

PRADO, F. & AGUILERA, A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *The EMBO Journal,* 24**,** 1267-1276.

RICHARDSON, S. M., LIU, S., BOEKE, J. D. & BADER, J. S. 2012. Design-A-Gene with GeneDesign. *Methods Mol Biol,* 852**,** 235-47.

RICHARDSON, S. M., MITCHELL, L. A., STRACQUADANIO, G., YANG, K., DYMOND, J. S., DICARLO, J. E., LEE, D., HUANG, C. L., CHANDRASEGARAN, S., CAI, Y., BOEKE, J. D. & BADER, J. S. 2017. Design of a synthetic yeast genome. *Science,* 355**,** 1040-1044.

RIO, D. C., ARES, M., JR., HANNON, G. J. & NILSEN, T. W. 2010. Polyacrylamide gel electrophoresis of RNA. *Cold Spring Harb Protoc,* 2010.

ROSSI, E., RIEGEL, M., MESSA, J., GIMELLI, S., MARASCHIO, P., CICCONE, R., STROPPI, M., RIVA, P., PERROTTA, C. S., MATTINA, T., MEMO, L., BAUMER, A., KUCINSKAS, V., CASTELLAN, C., SCHINZEL, A. & ZUFFARDI, O. 2008. Duplications in addition to terminal deletions are present in a proportion of ring chromosomes: clues to the mechanisms of formation. *J Med Genet,* 45**,** 147-54.

ROTHSTEIN, R., MICHEL, B. & GANGLOFF, S. 2000. Replication fork pausing and recombination or "gimme a break". *Genes Dev,* 14**,** 1-10.

RUIZ, J. F., GÓMEZ-GONZÁLEZ, B. & AGUILERA, A. 2011. AID Induces Double-Strand Breaks at Immunoglobulin Switch Regions and c-MYC Causing Chromosomal Translocations in Yeast THO Mutants. *PLOS Genetics,* 7**,** e1002009.

SANDELL, L. L. & ZAKIAN, V. A. 1992. Telomeric position effect in yeast. *Trends Cell Biol,* 2**,** 10-4.

SAUER, B. & MCDERMOTT, J. 2004. DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-c1 regions of P1-related phages. *Nucleic Acids Res,* 32**,** 6086-95.

SCHATTNER, P., BROOKS, A. N. & LOWE, T. M. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res,* 33**,** W686-9.

SCHRAMM, L. & HERNANDEZ, N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev,* 16**,** 2593-620.

SEKEDAT, M. D., FENYÖ, D., ROGERS, R. S., TACKETT, A. J., AITCHISON, J. D. & CHAIT, B. T. 2010. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Molecular Systems Biology,* 6**,** 353-353.

SHAMPAY, J., SZOSTAK, J. W. & BLACKBURN, E. H. 1984. DNA sequences of telomeres maintained in yeast. *Nature,* 310**,** 154-7.

SHEN, Y., STRACQUADANIO, G., WANG, Y., YANG, K., MITCHELL, L. A., XUE, Y., CAI, Y., CHEN, T., DYMOND, J. S., KANG, K., GONG, J., ZENG, X., ZHANG, Y., LI, Y., FENG, Q., XU, X., WANG, J., WANG, J., YANG, H., BOEKE, J. D. & BADER, J. S. 2016. SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res,* 26**,** 36-49.

SHEN, Y., WANG, Y., CHEN, T., GAO, F., GONG, J., ABRAMCZYK, D., WALKER, R., ZHAO, H., CHEN, S., LIU, W., LUO, Y., MÜLLER, C. A., PAUL-DUBOIS-TAINE, A., ALVER, B., STRACQUADANIO, G., MITCHELL, L. A., LUO, Z., FAN, Y., ZHOU, B., WEN, B., TAN, F., WANG, Y., ZI, J., XIE, Z., LI, B., YANG, K., RICHARDSON, S. M., JIANG, H., FRENCH, C. E., NIEDUSZYNSKI, C. A., KOSZUL, R., MARSTON, A. L., YUAN, Y., WANG, J., BADER, J. S., DAI, J., BOEKE, J. D., XU, X., CAI, Y. & YANG, H. 2017. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science,* 355**,** 1047.

SOARES, E. V. & SOARES, H. M. 2012. Bioremediation of industrial effluents containing heavy metals using brewing cells of *Saccharomyces cerevisiae* as a green technology: a review. *Environ Sci Pollut Res Int,* 19**,** 1066-83.

SOLLIER, J. & CIMPRICH, K. A. 2015. Breaking bad: R-loops and genome integrity. *Trends Cell Biol,* 25**,** 514-22.

SOLLIER, J., STORK, C. T., GARCIA-RUBIO, M. L., PAULSEN, R. D., AGUILERA, A. & CIMPRICH, K. A. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell,* 56**,** 777-85.

STAHMANN, K. P., REVUELTA, J. L. & SEULBERGER, H. 2000. Three biotechnical processes using *Ashbya gossypii*, *Candida famata*, or *Bacillus subtilis* compete with chemical riboflavin production. *Appl Microbiol Biotechnol,* 53**,** 509-16.

STEMMER, W. P., CRAMERI, A., HA, K. D., BRENNAN, T. M. & HEYNEKER, H. L. 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene,* 164**,** 49-53.

STEWART, G. G. & PRIEST, F. G. 2006. *Handbook of Brewing, Second Edition*, CRC Press.

STINCHCOMB, D. T., STRUHL, K. & DAVIS, R. W. 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature,* 282**,** 39-43.

STRABY, K. B. 1988. A yeast tRNAArg gene can act as promoter for a 5′ flank deficient, non-transcribable tRNA*SUP6* gene to produce biologically active suppressor tRNA. *Nucleic Acids Research,* 16**,** 2841-2857.

SUN, F. J. & CAETANO-ANOLLES, G. 2008. Evolutionary patterns in the sequence and structure of transfer RNA: early origins of archaea and viruses. *PLoS Comput Biol,* 4**,** e1000018.

SURRATT, C. K., LESNIKOWSKI, Z., SCHIFMAN, A. L., SCHMIDT, F. J. & HECHT, S. M. 1990. Construction and processing of transfer RNA precursor models. *J Biol Chem,* 265**,** 22506-12.

SZWEYKOWSKA-KULINSKA, Z., SENGER, B., KEITH, G., FASIOLO, F. & GROSJEAN, H. 1994. Intron-dependent formation of pseudouridines in the anticodon of *Saccharomyces cerevisiae* minor tRNA(Ile). *EMBO J,* 13**,** 4636-44.

TAKEUCHI, Y., HORIUCHI, T. & KOBAYASHI, T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes Dev,* 17**,** 1497-506.

THOMPSON, M., HAEUSLER, R. A., GOOD, P. D. & ENGELKE, D. R. 2003. Nucleolar clustering of dispersed tRNA genes. *Science,* 302**,** 1399-401.

TIPPER, D. J. 1973. Inhibition of yeast ribonucleic acid polymerases by thiolutin. *J Bacteriol,* 116**,** 245-56.
TROTTA, C. R., MIAO, F., ARN, E. A., STEVENS, S. W., HO, C. K., RAUHUT, R. & ABELSON, J. N. 1997. The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell,* 89**,** 849-58.
TUDURI, S., CRABBE, L., CONTI, C., TOURRIERE, H., HOLTGREVE-GREZ, H., JAUCH, A., PANTESCO, V., DE VOS, J., THOMAS, A., THEILLET, C., POMMIER, Y., TAZI, J., COQUELLE, A. & PASERO, P. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol,* 11**,** 1315-24.
TUGENDREICH, S., BASSETT, D. E., JR., MCKUSICK, V. A., BOGUSKI, M. S. & HIETER, P. 1994. Genes conserved in yeast and humans. *Hum Mol Genet,* 3 Spec No**,** 1509-17.
TUROWSKI, T. W., LESNIEWSKA, E., DELAN-FORINO, C., SAYOU, C., BOGUTA, M. & TOLLERVEY, D. 2016. Global analysis of transcriptionally engaged yeast RNA polymerase III reveals extended tRNA transcripts. *Genome Res,* 26**,** 933-44.
UEMURA, T., OHKURA, H., ADACHI, Y., MORINO, K., SHIOZAKI, K. & YANAGIDA, M. 1987. DNA topoisomerase II is required for condensation and separation of mitotic chromosomes in *S. pombe*. *Cell,* 50**,** 917-25.
VIJAYRAGHAVAN, S. & SCHWACHA, A. 2012. The eukaryotic Mcm2-7 replicative helicase. *Subcell Biochem,* 62**,** 113-34.
VORTLER, S. & MORL, M. 2010. tRNA-nucleotidyltransferases: highly unusual RNA polymerases with vital functions. *FEBS Lett,* 584**,** 297-302.
WAHBA, L., AMON, J. D., KOSHLAND, D. & VUICA-ROSS, M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell,* 44**,** 978-88.
WALKER, R. S. & CAI, Y. 2016. The Fifth Annual Sc2.0 and Synthetic Genomes Conference: Synthetic Genomes in High Gear. *ACS Synth Biol,* 5**,** 920-2.
WEI, Y. 2012. A Simple Preparation of RNA from Yeast by Hot Phenol for Northern Blot. *Bio-protocol,* 2(12): e209.
WEI, Y. 2013. Northern Blot of tRNA in Yeast. *Bio-protocol* 3(7): e464.
WELLINGER, R. E., PRADO, F. & AGUILERA, A. 2006. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol Cell Biol,* 26**,** 3327-34.
WENDLAND, J. & WALTHER, A. 2011. Genome evolution in the eremothecium clade of the Saccharomyces complex revealed by comparative genomics. *G3 (Bethesda),* 1**,** 539-48.
WILUSZ, J. E. 2015. Removing roadblocks to deep sequencing of modified RNAs. *Nat Methods,* 12**,** 821-2.
WU, Y., LI, B.-Z., ZHAO, M., MITCHELL, L. A., XIE, Z.-X., LIN, Q.-H., WANG, X., XIAO, W.-H., WANG, Y., ZHOU, X., LIU, H., LI, X., DING, M.-Z., LIU, D., ZHANG, L., LIU, B.-L., WU,

X.-L., LI, F.-F., DONG, X.-T., JIA, B., ZHANG, W.-Z., JIANG, G.-Z., LIU, Y., BAI, X., SONG, T.-Q., CHEN, Y., ZHOU, S.-J., ZHU, R.-Y., GAO, F., KUANG, Z., WANG, X., SHEN, M., YANG, K., STRACQUADANIO, G., RICHARDSON, S. M., LIN, Y., WANG, L., WALKER, R., LUO, Y., MA, P.-S., YANG, H., CAI, Y., DAI, J., BADER, J. S., BOEKE, J. D. & YUAN, Y.-J. 2017. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science,* 355**,** 1048.

XIE, Z.-X., LI, B.-Z., MITCHELL, L. A., WU, Y., QI, X., JIN, Z., JIA, B., WANG, X., ZENG, B.-X., LIU, H.-M., WU, X.-L., FENG, Q., ZHANG, W.-Z., LIU, W., DING, M.-Z., LI, X., ZHAO, G.-R., QIAO, J.-J., CHENG, J.-S., ZHAO, M., KUANG, Z., WANG, X., MARTIN, J. A., STRACQUADANIO, G., YANG, K., BAI, X., ZHAO, J., HU, M.-L., LIN, Q.-H., ZHANG, W.-Q., SHEN, M.-H., CHEN, S., SU, W., WANG, E.-X., GUO, R., ZHAI, F., GUO, X.-J., DU, H.-X., ZHU, J.-Q., SONG, T.-Q., DAI, J.-J., LI, F.-F., JIANG, G.-Z., HAN, S.-L., LIU, S.-Y., YU, Z.-C., YANG, X.-N., CHEN, K., HU, C., LI, D.-S., JIA, N., LIU, Y., WANG, L.-T., WANG, S., WEI, X.-T., FU, M.-Q., QU, L.-M., XIN, S.-Y., LIU, T., TIAN, K.-R., LI, X.-N., ZHANG, J.-H., SONG, L.-X., LIU, J.-G., LV, J.-F., XU, H., TAO, R., WANG, Y., ZHANG, T.-T., DENG, Y.-X., WANG, Y.-R., LI, T., YE, G.-X., XU, X.-R., XIA, Z.-B., ZHANG, W., YANG, S.-L., LIU, Y.-L., DING, W.-Q., LIU, Z.-N., ZHU, J.-Q., LIU, N.-Z., WALKER, R., LUO, Y., WANG, Y., SHEN, Y., YANG, H., CAI, Y., MA, P.-S., ZHANG, C.-T., BADER, J. S., BOEKE, J. D. & YUAN, Y.-J. 2017. “Perfect” designer chromosome V and behavior of a ring derivative. *Science,* 355**,** 1046.

YIP, M. Y. 2015. Autosomal ring chromosomes in human genetic disorders. *Transl Pediatr,* 4**,** 164-74.

YOO, C. J. & WOLIN, S. L. 1997. The Yeast La Protein Is Required for the 3′ Endonucleolytic Cleavage That Matures tRNA Precursors. *Cell,* 89**,** 393-402.

YOSHIHISA, T., YUNOKI-ESAKI, K., OHSHIMA, C., TANAKA, N. & ENDO, T. 2003. Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria. *Mol Biol Cell,* 14**,** 3266-79.

ZHANG, W., ZHAO, G., LUO, Z., LIN, Y., WANG, L., GUO, Y., WANG, A., JIANG, S., JIANG, Q., GONG, J., WANG, Y., HOU, S., HUANG, J., LI, T., QIN, Y., DONG, J., QIN, Q., ZHANG, J., ZOU, X., HE, X., ZHAO, L., XIAO, Y., XU, M., CHENG, E., HUANG, N., ZHOU, T., SHEN, Y., WALKER, R., LUO, Y., KUANG, Z., MITCHELL, L. A., YANG, K., RICHARDSON, S. M., WU, Y., LI, B.-Z., YUAN, Y.-J., YANG, H., LIN, J., CHEN, G.-Q., WU, Q., BADER, J. S., CAI, Y., BOEKE, J. D. & DAI, J. 2017. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science,* 355**,** 1049.

ZHENG, G., QIN, Y., CLARK, W. C., DAI, Q., YI, C., HE, C., LAMBOWITZ, A. M. & PAN, T. 2015. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods,* 12**,** 835-7.

ZIEHLER, W. A., DAY, J. J., FIERKE, C. A. & ENGELKE, D. R. 2000. Effects of 5' leader and 3' trailer structures on pre-tRNA processing by nuclear RNase P. *Biochemistry,* 39**,** 9909-16.

Appendix

Appendix I: Relevant Publications

The work undertaken in this thesis will form the body of a manuscript currently in preparation. Additionally, this work was presented at the 2016 5th Annual Sc2.0 and Synthetic Genomes Conference (Edinburgh, Scotland) and published in the following article:

WALKER, R. S. & CAI, Y. 2016. The Fifth Annual Sc2.0 and Synthetic Genomes Conference: Synthetic Genomes in High Gear. ACS Synth Biol, 5, 920-2.

The work undertaken during this PhD program of study also contributed to the following published articles:

MERCY, G., MOZZICONACCI, J., SCOLARI, V. F., YANG, K., ZHAO, G., THIERRY, A., LUO, Y., MITCHELL, L. A., SHEN, M., SHEN, Y., **WALKER, R.,** ZHANG, W., WU, Y., XIE, Z.-X., LUO, Z., CAI, Y., DAI, J., YANG, H., YUAN, Y.-J., BOEKE, J. D., BADER, J. S., MULLER, H. & KOSZUL, R. 2017. 3D organization of synthetic and scrambled chromosomes. *Science*, 355, 1050.

MITCHELL, L. A., WANG, A., STRACQUADANIO, G., KUANG, Z., WANG, X., YANG, K., RICHARDSON, S., MARTIN, J. A., ZHAO, Y., **WALKER, R.,** LUO, Y., DAI, H., DONG, K., TANG, Z., YANG, Y., CAI, Y., HEGUY, A., UEBERHEIDE, B., FENYÖ, D., DAI, J., BADER, J. S. & BOEKE, J. D. 2017. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science*, 355, 1045.

SHEN, Y., WANG, Y., CHEN, T., GAO, F., GONG, J., ABRAMCZYK, D., **WALKER, R.,** ZHAO, H., CHEN, S., LIU, W., LUO, Y., MÜLLER, C. A., PAUL-DUBOIS-TAINE, A., ALVER, B., STRACQUADANIO, G., MITCHELL, L. A., LUO, Z., FAN, Y., ZHOU, B., WEN, B., TAN, F., WANG, Y., ZI, J., XIE, Z., LI, B., YANG, K., RICHARDSON, S. M., JIANG, H., FRENCH, C. E., NIEDUSZYNSKI, C. A., KOSZUL, R., MARSTON, A. L., YUAN, Y., WANG, J., BADER, J. S., DAI, J., BOEKE, J. D., XU, X., CAI, Y. & YANG, H. 2017. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science*, 355, 1047.

WU, Y., LI, B.-Z., ZHAO, M., MITCHELL, L. A., XIE, Z.-X., LIN, Q.-H., WANG, X., XIAO, W.-H., WANG, Y., ZHOU, X., LIU, H., LI, X., DING, M.-Z., LIU, D., ZHANG, L., LIU, B.-L., WU, X.-L., LI, F.-F., DONG, X.-T., JIA, B., ZHANG, W.-Z., JIANG, G.-Z., LIU, Y., BAI, X., SONG, T.-Q., CHEN, Y., ZHOU, S.-J., ZHU, R.-Y., GAO, F., KUANG, Z., WANG,

X., SHEN, M., YANG, K., STRACQUADANIO, G., RICHARDSON, S. M., LIN, Y., WANG, L., **WALKER, R.,** LUO, Y., MA, P.-S., YANG, H., CAI, Y., DAI, J., BADER, J. S., BOEKE, J. D. & YUAN, Y.-J. 2017. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science*, 355, 1048.

XIE, Z.-X., LI, B.-Z., MITCHELL, L. A., WU, Y., QI, X., JIN, Z., JIA, B., WANG, X., ZENG, B.-X., LIU, H.-M., WU, X.-L., FENG, Q., ZHANG, W.-Z., LIU, W., DING, M.-Z., LI, X., ZHAO, G.-R., QIAO, J.-J., CHENG, J.-S., ZHAO, M., KUANG, Z., WANG, X., MARTIN, J. A., STRACQUADANIO, G., YANG, K., BAI, X., ZHAO, J., HU, M.-L., LIN, Q.-H., ZHANG, W.-Q., SHEN, M.-H., CHEN, S., SU, W., WANG, E.-X., GUO, R., ZHAI, F., GUO, X.-J., DU, H.-X., ZHU, J.-Q., SONG, T.-Q., DAI, J.-J., LI, F.-F., JIANG, G.-Z., HAN, S.-L., LIU, S.-Y., YU, Z.-C., YANG, X.-N., CHEN, K., HU, C., LI, D.-S., JIA, N., LIU, Y., WANG, L.-T., WANG, S., WEI, X.-T., FU, M.-Q., QU, L.-M., XIN, S.-Y., LIU, T., TIAN, K.-R., LI, X.-N., ZHANG, J.-H., SONG, L.-X., LIU, J.-G., LV, J.-F., XU, H., TAO, R., WANG, Y., ZHANG, T.-T., DENG, Y.-X., WANG, Y.-R., LI, T., YE, G.-X., XU, X.-R., XIA, Z.-B., ZHANG, W., YANG, S.-L., LIU, Y.-L., DING, W.-Q., LIU, Z.-N., ZHU, J.-Q., LIU, N.-Z., **WALKER, R.,** LUO, Y., WANG, Y., SHEN, Y., YANG, H., CAI, Y., MA, P.-S., ZHANG, C.-T., BADER, J. S., BOEKE, J. D. & YUAN, Y.-J. 2017. "Perfect" designer chromosome V and behavior of a ring derivative. *Science*, 355, 1046.

ZHANG, W., ZHAO, G., LUO, Z., LIN, Y., WANG, L., GUO, Y., WANG, A., JIANG, S., JIANG, Q., GONG, J., WANG, Y., HOU, S., HUANG, J., LI, T., QIN, Y., DONG, J., QIN, Q., ZHANG, J., ZOU, X., HE, X., ZHAO, L., XIAO, Y., XU, M., CHENG, E., HUANG, N., ZHOU, T., SHEN, Y., **WALKER, R.,** LUO, Y., KUANG, Z., MITCHELL, L. A., YANG, K., RICHARDSON, S. M., WU, Y., LI, B.-Z., YUAN, Y.-J., YANG, H., LIN, J., CHEN, G.-Q., WU, Q., BADER, J. S., CAI, Y., BOEKE, J. D. & DAI, J. 2017. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science*, 355, 1049.

Appendix II: Genome Accession Numbers

Table I: Genome Accession Numbers. The table below describes the accession numbers used to recover genome sequence files.

Chromosome	*S. cerevisiae*	*A. gossypii*	*E. cymbalariae*
1	NC_001133.9	NC_005782.2	NC_016449.1
2	NC_001134.8	NC_005783.5	NC_016450.1
3	NC_001135.5	NC_005784.3	NC_016451.1
4	NC_001136.10	NC_005785.6	NC_016452.1
5	NC_001137.3	NC_005786.2	NC_016453.1
6	NC_001138.5	NC_005787.5	NC_016454.1
7	NC_001139.9	NC_005788.4	NC_016455.1
8	NC_001140.6	-	NC_016456.1
9	NC_001141.2	-	-
10	NC_001142.9	-	-
11	NC_001143.9	-	-
12	NC_001144.5	-	-
13	NC_001145.3	-	-
14	NC_001146.8	-	-
15	NC_001147.6	-	-
16	NC_001148.4	-	-

Appendix III: List of PCR Tags for Neochromosome Verification

Table II: List of PCR tag primer sequences used for verifying the presence of each tRNA cassette. "Well" refers to the corresponding well on the 96-well agarose gel, and "tRNA ID" refers to the naming convention used for each tRNA gene (see Appendix II, **Table III** for the corresponding gene name of each).

Plate	Well	tRNA ID	Fwd Primer	Reverse Primer
Plate 1	A1	SC.t06.01	GTGATTGGACCGCTAGAGC	CCATCTAAGGGTACCAAAAAATGG
Plate 1	A2	SC.t06.02	AGTGGTATCGATTCTTTCCTACTG	gttaTGAAGGTACCACAAAAATAACG
Plate 1	A3	SC.t06.03	CATTACAGAAGCAATCTTTTGTAAGG	aTTTACCATGTGATTCAAAAATGCG
Plate 1	A4	SC.t06.04	GCACACCCAAAAGACATTCG	CGTTTCGCCATGAAAAAAAAATCTC
Plate 1	A5	SC.t06.05	AACTGGCTGCCTTATCTTGC	TGGAGACAGACGAGACATAGG
Plate 1	A6	SC.t06.06	AGGATTTGGAGCCTTACATATTCG	ATACCGCGAGCAAAAAGATG
Plate 1	A7	SC.t06.07	GCACAACTACTGGCGTACC	aCCAGAGTATCCAAAAAATAACGTC
Plate 1	A8	SC.t06.08	ACGAGATCAGCAACTTTTTGAGT	gttaTCATAAAAACATTAAAAATTTGGACG
Plate 1	A9	SC.t06.09	GTGCCTTTACAAATCAAATGACG	GGTTCGCTTACCAAAAAAAATCTCC
Plate 1	A10	SC.t06.10	CTGTGACCGCGATAAAAACG	GAAGCGTCGCATAAAAATTAAGC
Plate 1	A11	SC.t05.01	tctcgcttatgtcatcaagagag	gtgctcgcgtcaaaagc
Plate 1	A12	SC.t05.02	gcaccgttatgcagtctctg	gggtgatatagcggttgttac
Plate 1	B1	SC.t05.03	taccgaacttcgcaagacc	gtccagctccaaatacttatttccc
Plate 1	B2	SC.t05.04	agtatattactgcaatccatttggc	cgaatgcacattgtgccc
Plate 1	B3	SC.t05.05	tactcgtgtcgaagaccgc	gggcagataaaaaatttcacgttc
Plate 1	B4	SC.t05.06	atcctccgtgtagagtagagg	gcgattagttaaaaaggtgcgtag
Plate 1	B5	SC.t05.07	gattggttttgaccggatgg	aatgtcaactagaaaaaatactccg
Plate 1	B6	SC.t05.08	gtgcttttgtacaagaccagc	gtgatagcctcgttccagtg
Plate 1	B7	SC.t05.09	atttgaagaacctcctcagcg	cacgtgatcaaacggaaactaac
Plate 1	B8	SC.t05.10	atatgctccagcgggtatg	gagtttcatctaaaaagaaacggc
Plate 1	B9	SC.t05.11	gactcctctgatacaatgcacc	gcaaccattgcagttgaaatac
Plate 1	B10	SC.t05.12	ctgcatggagtatccagaaaagg	aagttatgccactgtctttccc
Plate 1	B11	SC.t05.13	caagttgccagatttgaaagc	aatacaaaaacagcaaaaaaatgcc
Plate 1	B12	SC.t05.14	tggagggtatggtttgtacg	tgaagaaatgcaaaaaagttaagcc
Plate 1	C1	SC.t05.15	gggcatctccgaaattttcg	acatgtctttagcggaaagc
Plate 1	C2	SC.t05.16	ccatgtatatttgagagccacag	tctaatcacgtttgtacccagg
Plate 1	C3	SC.t05.17	ccgagctcaagatctggg	acagatatataatggtgatactgcattc
Plate 1	C4	SC.t05.18	accgattttataatcactgtgtagc	tgtaactaaactgccaagtcaaatag
Plate 1	C5	SC.t05.19	cgtacgtcatcatctttacgtcg	caaaaatcgacaaaaggcgg
Plate 1	C6	SC.t05.20	gtctctaagccgaacataaatagtc	cgaagaactcaaaagatcagcc
Plate 1	C7	SC.t04.01	gttagtctcgatagccacacag	cgggatcaccttctgtaacc
Plate 1	C8	SC.t04.02	gctttttacatagacttcggcg	aggtggaatacggtagaaaagg
Plate 1	C9	SC.t04.03	ttaggtgactgaggcccac	gctgatgtgtccagagaaattg
Plate 1	C10	SC.t04.04	cactaactagcctgtgtaccc	ccttcgtgatttcaagcttgg
Plate 1	C11	SC.t04.05	gccaattatttaaagttatacttgtcgagtc	cgatgtatatacagattagactgaacagc
Plate 1	C12	SC.t04.06	gcctgtatcgagattttcctatagg	accccataaaagtttgaccgg

[Table continued on following page]

Plate	Well	tRNA ID	Fwd Primer	Reverse Primer
Plate 1	D1	SC.t04.07	cttgtcgagtcataatacaaaaaagg	ttgctgataacaatacatcgaatcg
Plate 1	D2	SC.t04.08	gcaaacaagctttgcatataatgc	ctccggtagactttgtcatcc
Plate 1	D3	SC.t04.09	accgtctcctcaacaaaaattag	tgctaaataaggacatcgatccc
Plate 1	D4	SC.t04.10	tcacgtgactccaattgcg	gaaatgagctacttgagacgc
Plate 1	D5	SC.t04.11	gacaaaaagctgaaacaagaggg	tcgatgttgcttgcaaaagg
Plate 1	D6	SC.t04.12	gtttcgaatcatgctcttccc	acgtacggtagtgactacctac
Plate 1	D7	SC.t04.13	gcaattgtgtcttttggccc	attgctgcccaaactttgg
Plate 1	D8	SC.t04.14	tttggcttttgagctttgctc	tgggttgtttcaaggttggg
Plate 1	D9	SC.t04.15	cggatataaatctttcctggggg	aagcatgcttactcagagagag
Plate 1	D10	SC.t04.16	gtaattagcacaaaggctggc	cgtcttgaaggacagtgcc
Plate 1	D11	SC.t04.17	gcaccaaaatgccagcaag	ggcttccggattattctgagg
Plate 1	D12	SC.t04.18	ccatttctaccactgcaatatgc	aacgatatcttggttagcagcc
Plate 1	E1	SC.t04.19	cgccaaaaaaatttcaaatcagaag	gctatggatagatgctaccgg
Plate 1	E2	SC.t04.20	gagaaaaccagtgcaaaaaagaatc	ggttttccttcctgctcgg
Plate 1	E3	SC.t04.21	tgaagatccctaacgctggg	gcgatttcagcctttattccttg
Plate 1	E4	SC.t04.22	ccagaaagaaaaaggacaaaagcg	aagggaattctctagacgtatgg
Plate 1	E5	SC.t04.23	ccttttgaagtaccaaaattacacg	tacaatgatcttattcctgccac
Plate 1	E6	SC.t04.24	caaacgcgaagaccaaaaaag	gatgttctagatctgtgccgg
Plate 1	E7	SC.t04.25	aataaaatatttagcaaaaaagatgcgg	gctatagtctagttggtttcagcc
Plate 1	E8	SC.t04.26	taacaatgcatgtgtatgctgtgg	tattatgccatgtctaccttgtagg
Plate 1	E9	SC.t04.27	gggtgtttgtttatcggctc	cttgttccccaggtagttcg
Plate 1	E10	SC.t04.28	gaattgatattgcattggaactcc	agtaaaataatgattccgcttcgtc
Plate 1	E11	**ARS1**	GCATTTGCGCACACTTTATAAATTC	GCCAACATTAGATAATGATCTGGATG
Plate 1	E12	SC.t03.01	tcagaaagtttctgcaaaaaataactc	gaagacctgtattggtaagtgc
Plate 1	F1	SC.t03.02	cgtgatccatgcaaaaaataatgg	aagacatagagtaacatctacctgg
Plate 1	F2	SC.t03.03	aaaaaaaagccacaaagggc	tggcgtagtattggattctcg
Plate 1	F3	SC.t03.04	ttgtcatacatagcacaggaaatg	gttagtagccattgcactatggg
Plate 1	F4	SC.t03.05	acatgtcttccacagcacg	gtcttcatttaacgctgaatggg
Plate 1	F5	SC.t03.06	tttaaagttagtgcgctggcg	gcacacagtagcttctgaaacc
Plate 1	F6	SC.t03.07	agacagcgagcggtttag	tttgatctttacactgaacgcg
Plate 1	F7	SC.t03.08	tagctaaactgcaccggaaag	gtggttttttccacgtactcg
Plate 1	F8	SC.t03.09	gcagaggaacaaaaagatgagc	gctgtgttttaagtgcagagc
Plate 1	F9	SC.t03.10	ccaaaaagactgcatgaggc	cttctaggtacggaacaagagg
Plate 1	F10	SC.t02.01	GCATTCTAATCTAGAAAAAAAAGTACTGAAG	AGGTATTGTCCATGCAGTACC
Plate 1	F11	SC.t02.02	GTGACGTGAGCAAAAAATATGC	GAGCTAGATCGCTCATTATTTTTGG
Plate 1	F12	SC.t02.03	ACCACCTCCAAAAAATATACGCC	GCCCACTGTCAGGATTATTCC
Plate 1	G1	SC.t02.04	CGGATCGCCAATCACCTC	TACTATACAAGCCAGCAGATTATGC
Plate 1	G2	SC.t02.05	CCTGAAACAGCATCAGATGC	AGATACGCATTGCTCGTATCC
Plate 1	G3	SC.t02.06	CGTCGGTTAAACAATGACACG	TTGACTCATTTATTACCTCAGAGGG
Plate 1	G4	SC.t02.07	ACTCAGTGCACCAAAACTGG	CTCAAAGGACAAAACAACGGC
Plate 1	G5	SC.t02.08	GAAAGCAAAAACTTCAAAACATTGG	GGAGCAAGGGTTATATAGAGCG
Plate 1	G6	SC.t02.09	CGATTTATGCTCACTTGCATCC	GGATCGTAGGCGGCAATC
Plate 1	G7	SC.t02.10	GGAGATAATTAAGCAGAAAAAAAAAATTAACAC	TCCACAAGCAAGGAAATCCG
Plate 1	G8	SC.t02.11	CGACAACCGTTTCAATAAATGTCAC	GCTTCAGGTCAACAAAATAATCTTG
Plate 1	G9	SC.t02.12	TAAAACCCGTCACCTCCAAC	CCCCATATTACAGACAAATACACC
Plate 1	G10	SC.t02.13	GCGCCTGGATGAACAAAAATTTTAG	TGATTTGGTACAGCTCCTTGG

[Table continued on following page]

Plate	Well	tRNA ID	Fwd Primer	Reverse Primer
Plate 1	G11	SC.t01.01	cgcagactaaagatacgatagcc	ccaatgcaggttgaaaagttgg
Plate 1	G12	SC.t01.02	gatagcatcaacaataaatgatttgc	gcaaagcacttaacatctcagg
Plate 2	H1	SC.t01.03	ggcaatcagatgacaaaaaatcatc	agtatttcaaccagcttcacaag
Plate 2	H2	SC.t01.04	cgtgaccatgtatacatggattcg	tgatatcctccgagtccatgag
Plate 2	H3	SC.t07.01	gctctgaaaaagttgacagtatagc	cttccttggtaggtctactaattcc
Plate 2	H4	SC.t07.02	gttcccgcagaaatcaaaaaaatg	acgaaaggcccatttttgattatg
Plate 2	H5	SC.t07.03	ctagctttcaatgcaaaaagagaac	tccactcttccaagctggac
Plate 2	H6	SC.t07.04	gtttgctgagctaaaaacaaaaaag	ttcaagggagaattgacaaatcg
Plate 2	H7	SC.t07.05	aatcattagtaatcaaaacggaaatagg	ggcatctatgctatgccatgtc
Plate 2	H8	SC.t07.06	acgcccgcaaaaatattgg	gcaccatcgttcgagatgg
Plate 2	H9	SC.t07.07	acgagatcaaatgctgacgtg	ctctaatttcagtgtttcctgcc
Plate 2	H10	SC.t07.08	gcaacttcactctaatgagagagtg	gggatgatccagttgcttgg
Plate 2	H11	SC.t07.09	catgaaaattccaaaaaaacagtactcc	gaagaaagaagaaagggtctaaacc
Plate 2	H12	SC.t07.10	tccaaaatgtcaaaaatcaaaaaaataaatg	cgtaggtaaagccgtgtttc
Plate 2	A1	SC.t07.11	gatctgagatgttctattccagaac	ccgtttaccgtagcttccg
Plate 2	A2	SC.t07.12	tattcgttgtatggcaacaaaaaag	ctaacattcttataatcgtctaacaactc
Plate 2	A3	SC.t07.13	gaagtacacgcagtcaaaaatattg	cttccttgttgctcctaccg
Plate 2	A4	SC.t07.14	tctaaaaattttcaacagcccagg	gttcaacaacgttatgattgagacc
Plate 2	A5	SC.t07.15	acgcaccctagtaacaaaaaac	tccactatgtaaatattgcaaccag
Plate 2	A6	SC.t07.16	gataaagccggcgaaatcc	aggcattgtcgtctttccc
Plate 2	A7	SC.t07.17	ctggtagacaagtaccaaaaaagg	cgctgatatcaagtctttgattgag
Plate 2	A8	SC.t07.18	cctatataagtaggaaaatgaaaaaaaacgac	tggttaatgtgatcggtgcc
Plate 2	A9	SC.t07.19	agcagtgcaagaaaaaagtaagac	gctgctcagaaaggaacgg
Plate 2	A10	SC.t07.20	gtcgctgctggaaaaaatcag	acgctattaaggatccacatcc
Plate 2	A11	SC.t07.21	gtaaagcgtacaacaaaaaaattgg	caaaataataaatattaatcaagataaagttaatagtgg
Plate 2	A12	SC.t07.22	agcataaatattcaatactacccgc	gcgctcgaaatagcatctcc
Plate 2	B1	SC.t07.23	agatctatgcacgggagcc	accaacagtttgcaccgatg
Plate 2	B2	SC.t07.24	agactggcttcaatcaaattagg	ttagttgattcaacatttcagcgg
Plate 2	B3	SC.t07.25	agtgagctttctcattttcaatgag	gcccgcggcagagag
Plate 2	B4	SC.t07.26	cgggcgtatgctgaaaaaaac	gctttagtttaggctcctcgc
Plate 2	B5	SC.t07.27	ttcttccgcgggactcttag	cgtgacagtaatgattgtccatacg
Plate 2	B6	SC.t07.28	tgcggcattctgaatctgc	tgtcgctccattcttcacg
Plate 2	B7	SC.t07.29	cgtgagtgatcatgatgggtac	atgctcgaaaaggtttcacg
Plate 2	B8	SC.t07.30	cttacctgagcagacttaatttgg	acctggatctgcgcaatac
Plate 2	B9	SC.t07.31	aacttttcacagacttagaagaagc	cggactctttgatcgaagtatcc
Plate 2	B10	SC.t07.32	ggcaatccttttttcaatggagc	gtggaaacaaatagaccaattttgg
Plate 2	B11	SC.t07.33	taaagttagtgcaagcgatgcc	ggttttcagagcttgctttacc
Plate 2	B12	SC.t07.34	gctgaacgcgcaaaaaattaaatg	gtggcaaagcattttgttgc
Plate 2	C1	SC.t07.35	gcacttagctatatcggtttagttg	ctatcactgttgtcaaccttgc
Plate 2	C2	SC.t07.36	ggcgctatttatttacttttcttcg	ttcactcctattaaccagcttcac
Plate 2	C3	ARS2	TTGGTCCACTGATAGTATCGC	TGGCTGAGATCTCAACTAATCG
Plate 2	C4	SC.t16.01	gcattagtcgccacaccaag	gttaccatctatatggtatttttatatatatgattcaaaaag
Plate 2	C5	SC.t16.02	ggtcgcgatcatcttcaaacc	tgtgtcctcgttacatgttaaaaaatg
Plate 2	C6	SC.t16.03	atcttcatcctagtaatatcatccacc	gttacaacatacatagaatacatacatagaatacag
Plate 2	C7	SC.t16.04	aactgcctacgtacaagaagc	cgtaagtgcgaataaatagcctttc
Plate 2	C8	SC.t16.05	atacttgaatagtgcgaaaagtgtc	tgtttcatgcgccaaattgag

[Table continued on following page]

Plate	Well	tRNA ID	Fwd Primer	Reverse Primer
Plate 2	C9	SC.t16.06	ttagtactgtgtctcagcaactg	tctcgagaacttttcacgcc
Plate 2	C10	SC.t16.07	cacgttacccacacaaaagaataac	cattatctacgagcatctcagc
Plate 2	C11	SC.t16.08	tttcccaagacaaaccatttctacc	gatgtatcttctaaaaacataaacgtcag
Plate 2	C12	SC.t16.09	agttttaggattatcaagttgtggc	ccgtagtcatatgataatatgagttctc
Plate 2	D1	SC.t16.10	tcacttcaatcttcttcatcgaagg	caaaaaccgcaggttcgag
Plate 2	D2	SC.t16.11	ggaggggactgttaaatactttcag	catacaggggtattagtttcttcaac
Plate 2	D3	SC.t16.12	tcgagttgtactgtagaaattggc	gtctgcgtagaggaaagatattctag
Plate 2	D4	SC.t16.13	caccattattcatcgaaataatgcc	tgtaaaaaattggctattgttcacc
Plate 2	D5	SC.t16.14	tgcttcttagcttggttatccatg	gcgtggaacgttatcaagg
Plate 2	D6	SC.t16.15	tgactctttatttgcttgttacgtc	aaagttactagagaaacccgcc
Plate 2	D7	SC.t16.16	tttctacttctctatattgctagggtg	ggagataagtaaaaaagagtgcgc
Plate 2	D8	SC.t16.17	taaatgccgctcttgagactc	gaatgttacttttcgcaaacaactg
Plate 2	D9	SC.t09.01	ccaggcacatccatacatcc	acaacgcaaaattgttctaactagg
Plate 2	D10	SC.t09.02	cgaaaccgcgaaagttatcg	tcgcacaagaagatcgagaag
Plate 2	D11	SC.t09.03	ggcgccgcaaaaaaactc	cgcagagtttccgaaacgg
Plate 2	D12	SC.t09.04	cccgggtagcatataaaaaaatgg	gcacttgcagataagtgagctg
Plate 2	E1	SC.t09.05	cgcaacaataaaaataacgacacc	aggggacgaaaatatgcacc
Plate 2	E2	SC.t09.06	acaaaactattcgctgaaaaaagcc	cgtcatctgtaatgctttttcgag
Plate 2	E3	SC.t09.07	cttgagccaaaaaagctccg	ctttggttgcattcatgtatgagg
Plate 2	E4	SC.t09.08	aactgctctcatgccaaagc	gtctatggaagcgtgcttcg
Plate 2	E5	SC.t09.09	agcctcgggcaaaaaaaac	cctccttttcttgcaacacg
Plate 2	E6	SC.t09.10	tgttagtgtgagaagtccaagc	attcccgagctcacatattcc
Plate 2	E7	SC.t08.01	caaaaaaactgcaaaaggaaatgc	gcacgtatttctccttcacacc
Plate 2	E8	SC.t08.02	aaccaaaagaacgccctttttaac	atacttctaaagggcaattcaacgg
Plate 2	E9	SC.t08.03	agtttcaaaaaatatcacgtataaaatacttgc	tatttggaaaacgtcaagtcgc
Plate 2	E10	SC.t08.04	gcactgttaaacaacaattcgttg	aagtatgagatgaatcgaatatcgg
Plate 2	E11	SC.t08.05	accttttttatgtaattttatggaaaaattagg	aaaaactcttacggtaacccag
Plate 2	E12	SC.t08.06	gccggtgggatcaaaaaattg	caggctttcctgtagcttcc
Plate 2	F1	SC.t08.07	catttagttcggaaggggc	cattaggctcttcgtttaagagc
Plate 2	F2	SC.t08.08	gttgggtggaaccaaaaaaagg	gaaattgcgtcactgtcacc
Plate 2	F3	SC.t08.09	attcctactcgaaagccgc	cgccgatcttagatttcaagg
Plate 2	F4	SC.t08.10	gcagacttttcatcgaacgc	gcgattgacatcttcaccgg
Plate 2	F5	SC.t08.11	taacatgacatcacgtgacagtc	ccagccctttgcgtactg
Plate 2	F6	SC.t10.01	cacgtgactagttgacttgatacg	gtcacgtggtgaaacaagtag
Plate 2	F7	SC.t10.02	aaggggaatggaaactagaacag	aattgatcaggaaagaagaagagg
Plate 2	F8	SC.t10.03	agttagaatcgaaccagcgac	aagttagtttaaatgcagctggg
Plate 2	F9	SC.t10.04	gctcatcgctagacgtaaatacc	ataaccactgcagagtaaattggc
Plate 2	F10	SC.t10.05	acctcttaaaatattcgtaatagacagc	aacgaattaaacctctctccatg
Plate 2	F11	SC.t10.06	cattgcaaagcgtctggag	cgttgtttcgtacatctctgg
Plate 2	F12	SC.t10.07	tgaaaacgttatcatggcattaattc	gtctttttgaaatagccacaatgc
Plate 2	G1	SC.t10.08	tacggttgtaactgttttgagactc	tccacacatcataatgtactacagg
Plate 2	G2	SC.t10.09	caacagatttttttcgaccagc	tcagaagagtaagaattaacttcgc
Plate 2	G3	SC.t10.10	gttgtatgcaatgccttcgg	gtgctactgccaggagatac
Plate 2	G4	SC.t10.11	gtcagcgcgatatttaaccg	actgcgaactacgagcatg
Plate 2	G5	SC.t10.12	cgctcctgtccaaaaaatatgc	ccgatggataccatatccagg
Plate 2	G6	SC.t10.13	ggctaagttttctggtgagaatacc	catttgtgaaccacactagcc

[Table continued on following page]

Plate	Well	tRNA ID	Fwd Primer	Reverse Primer
Plate 2	G7	SC.t10.14	atcagaatttttcaacaatttccacg	gcctatacgtaaaccaagactcc
Plate 2	G8	SC.t10.15	cagtcttcgagttggtctgg	cttgctgaagatcaagcatgg
Plate 2	G9	SC.t10.16	agtacaactcaacgtgagctc	cgccctcatgtgttctacc
Plate 2	G10	SC.t10.17	ggaaatcgatccgaagtttgttattatc	cgaatcgcagcaatagatttttg
Plate 2	G11	SC.t10.18	cgcagcagggaaaaaagtag	cattcagtgtcccgttccac
Plate 2	G12	SC.t10.19	tttcatcgaagcgagggtg	gtggtcgtacacattcttgtg
Plate 2	H1	SC.t10.20	aaaatgggcgaaattttaaatccg	acttcgtaaacttgttgagaatcc
Plate 2	H2	SC.t10.21	ttgaaatgtttttggccactagac	tagtatccaaaattcccagtgcag
Plate 2	H3	SC.t10.22	gcgcaaaaaagcaaaaacaac	gcgtcgctactaaatgccag
Plate 2	H4	SC.t10.23	ggcaacattgtagaacaggaac	gagaaaaagcagcggttagc
Plate 2	H5	SC.t10.24	caattatttaaagttactatgctttattatgactg	gttaatatacgtcattgttctgcgc
Plate 2	H6	SC.t11.01	aaaaccgccaaaaataaagtgc	cagtatatgctgactttacgatgg
Plate 2	H7	SC.t11.02	gcgatgtacatgtacaaaaaaaagg	caggtgcggtaaacctatgg
Plate 2	H8	SC.t11.03	tagtgaccaagatattactatactgcaaaaag	tacagttccgatgattgtttcacc
Plate 2	H9	SC.t11.04	ctcgagagaaaaattacaatagaactgg	gatcatatctgatgcttttggtgg
Plate 2	H10	SC.t11.05	taacccattttcaaacccctgag	cctatgtatttgacatcattcgagg
Plate 2	H11	SC.t11.06	gtcaggctgaaaaggaaaaaagaga	gtttttcagctttgaagagggc
Plate 2	H12	SC.t11.07	cagtctaagctgcaaaaaaattgg	cgtgtctataatttgcttgtagagg
Plate 3	A1	SC.t11.08	ggatgtggaagtttcaaacgg	gtcggatcgacagatttttgc
Plate 3	A2	SC.t11.09	taaactagctgtgaaccatggtttg	gcctccactatttccaatttgc
Plate 3	A3	SC.t11.10	gttagtttccgatttaacaagccc	acccaggagccaaaaaacc
Plate 3	A4	SC.t11.11	ccgcgcgatactttagcc	agtctcgcgagaaccagg
Plate 3	A5	SC.t11.12	gtaacattgttaacacaaaaaagaatactgg	aagagttgtatgattaggatgaacg
Plate 3	A6	SC.t11.13	agttatcataaatcaagggaataagagc	tcccaatatcaccagaatactaacg
Plate 3	A7	SC.t11.14	gttaatacttggatgttcacacaaatatc	actcatagaaatactgcagagtcc
Plate 3	A8	SC.t11.15	tatttaaagttagcccgtgcacc	caatgcgaggcaactaaacc
Plate 3	A9	SC.t11.16	aatttcggaggtttgtgtagg	ttaaatcgaacccatccacg
Plate 3	A10	SC.t13.01	caatgcgaacattgaaatgtgag	gttagtcagaacaaggtgctgg
Plate 3	A11	SC.t13.02	taaagttaaaagactggatttggcc	tatatattacagcctgtcactgctc
Plate 3	A12	SC.t13.03	tttattcatcagcaccaggttgc	cccaatggttaataccacctgtg
Plate 3	B1	SC.t13.04	aatcagatgcgggtcaaagtc	tgggcctgttgataagttctg
Plate 3	B2	SC.t13.05	atcacaagaatgaaaaaagaacgac	tgattcaaacttttgaaccctcc
Plate 3	B3	SC.t13.06	tctggctttagcatccctcc	gggtttggttcaccaactcc
Plate 3	B4	SC.t13.07	aacgtccaattaatagtctgcacc	cggcactgcaaatattttcgg
Plate 3	B5	SC.t13.08	atggaagttaacggcaaacataac	gcgaaaatgtgtgtgaactgg
Plate 3	B6	SC.t13.09	cccagtttaatcatcatgttaaaatcac	gcagcatatttgcaaacttttgg
Plate 3	B7	SC.t13.10	cctcaacatgtgatgctaaagg	tgatttctagaatttgttgaactaaaaaagg
Plate 3	B8	SC.t13.11	accaaattaccaattttgtaggctag	ttcatatctatgacgcatggtcatc
Plate 3	B9	SC.t13.12	cttctaataccacaaagccagc	ataacttttgtgacctttactaggc
Plate 3	B10	SC.t13.13	attttacgtgtatgctatatatacacagc	tggtatatgtcgactataatgtcatcg
Plate 3	B11	SC.t13.14	gtaaaaagaattcgaaaagtgaagcc	tctttactccaaatttgctagaagcc
Plate 3	B12	SC.t13.15	gcctgtctttttccctctcc	ggagcagaagaacaccaagc
Plate 3	C1	SC.t13.16	atagccaactgctctggc	ccaaataaatgccttgatctcgg
Plate 3	C2	SC.t13.17	gacaagcccgaaaatgcc	atccgaggtatgatacgtgact
Plate 3	C3	SC.t13.18	ggctgcggaaaaaataacactc	gctttttgtttccctgacgc
Plate 3	C4	SC.t13.19	atgtcaaaaagcagaaaactactgg	gtattgatgtgatgtgaccgaatc
Plate 3	C5	SC.t13.20	ggggaaaaaattacggagttcag	ccaccgttaaaacgaaggtg

[Table continued on following page]

Plate	Well	tRNA ID	Fwd Primer	Reverse Primer
Plate 3	C6	SC.t13.21	ccgtgtggcaaaaagtctc	ctggtctggcagttcaaatg
Plate 3	C7	SC.t12.01	gatattcacctttcccgcgg	cccgcatattatatatgcaaacccc
Plate 3	C8	SC.t12.02	gcgcaagcaaaaaataacgac	ggatcatttgtccttgcagc
Plate 3	C9	SC.t12.03	caaatcgggtggtctattcaac	gcatatttcttttcgccgtctc
Plate 3	C10	SC.t12.04	acacgtgatcacaatgtgtttg	gccaactttgttccgtataaacc
Plate 3	C11	SC.t12.05	ccgtcattgaacttttgaaaattcg	ccacaagtgtagatttctgaaggg
Plate 3	C12	SC.t12.06	gattgctccttcttagttgtaaaagag	gttgaggaacttgtcgttcag
Plate 3	D1	SC.t12.07	cttggtctacagcgtgctg	aaggaccactaggccagc
Plate 3	D2	SC.t12.08	atcttctacctttttagtgtggtcc	gccaatcttgcattccttttgtg
Plate 3	D3	SC.t12.09	gaagacaaaagcttggcaaaaag	gctagcttacgctttagttgag
Plate 3	D4	SC.t12.10	aaactacagtcaaactatatggacg	gtggaaaggttctttttcatcattc
Plate 3	D5	SC.t12.11	tttaaagttatctgcaggcggg	gaagctgtatctaagaagagtgagg
Plate 3	D6	SC.t12.12	gttagccgcaagtggagag	gtgatttcctggccagacg
Plate 3	D7	SC.t12.13	tgcagacatgtggcaagag	ctgtcatccttgttggcgg
Plate 3	D8	SC.t12.14	tatccggacactcaaatcaaacttc	agcgactcatttgaaatgacctaag
Plate 3	D9	SC.t12.15	ccgtatgagatctaaacaaattgcg	cagttggattacgaatcaacacg
Plate 3	D10	SC.t12.16	gtaaaaataaaaattatgtcagaatcctccg	actctcagtgtttgtgtttatctgg
Plate 3	D11	SC.t12.17	ctaccagcgccaaaaagaag	ccctgttgaatatctgtgtgtc
Plate 3	D12	SC.t12.18	gttagtaatcaaacggttatccatatgtg	ggctgtttttttcttttctgtttgtg
Plate 3	E1	SC.t12.19	ctttgccagaagaactgtaccc	cagttctactgggaatctgtaagag
Plate 3	E2	SC.t12.20	gttgtcttaaagatacaaaaacgacc	cgacgctcttaattttcaatgg
Plate 3	E3	SC.t12.21	tgtttatgcatcctgaatgcatg	ccgtaattcaaagagaatgtaaatgg
Plate 3	E4	SC.t14.01	gccaggttcttcttgcaaac	caacttgcttctttggtggc
Plate 3	E5	SC.t14.02	acattttgggtgccatttgg	gcttctccgttttacgatttgtc
Plate 3	E6	SC.t14.03	actttaaacgaatttgaaaaatgaatttgc	agggcactctacattttagaaagg
Plate 3	E7	SC.t14.04	tgtcattgttattttatgtcaaccgtg	ctcataattactacgataaacagtgtgc
Plate 3	E8	SC.t14.05	aagttacccatggacttgtgc	cggtctcatccatagctttgc
Plate 3	E9	SC.t14.06	accattgaactgtgtctgcc	cgatatgcaggtttcaagctc
Plate 3	E10	SC.t14.07	ctgcatggacagatcgtcag	aagaagattgtcttgtgtgcttg
Plate 3	E11	SC.t14.08	cttccgtgacattcggtctg	cttgattgtgctctgttgagg
Plate 3	E12	SC.t14.09	cagggaaaagaaaaacaaagtgaac	cgtcttgaacgttccgatataattg
Plate 3	F1	SC.t14.10	gcgatagaataatgtgctagtatgac	tctaaaggatcagttccctctgg
Plate 3	F2	SC.t14.11	gacaaaaataagcacattggcctc	aaatctgtagaacgacgacacc
Plate 3	F3	SC.t14.12	gcccgacaaaaagcttgaag	ggcaaaacgtacttcatttcacc
Plate 3	F4	SC.t14.13	cccaaatgtcgatatgaaatgcc	agtcgatcaagatgctaatcttcg
Plate 3	F5	SC.t14.14	agttacacgtgatatacgcgtg	agcacccactacctataaccg
Plate 3	F6	SC.t15.01	cccaaattacggcataaatgaaacc	agctagcttgggtctccttatatac
Plate 3	F7	SC.t15.02	gttgtgtaaattgggaagtgttgc	ctcttcaacgacgcacttaag
Plate 3	F8	SC.t15.03	ctacatatacataccctcctgcc	atgtaatacgtcatgttagagcgg
Plate 3	F9	SC.t15.04	ccgggaattgtaaaaagatattccg	gccaattactacgattacttatcgg
Plate 3	F10	SC.t15.05	taaagttattcagcactcgcacc	tgattaagctctaagtgttgcgg
Plate 3	F11	SC.t15.06	atatccagaaacctcttccaaaaagag	taggtggtcttacttatcatactcaatg
Plate 3	F12	SC.t15.07	tgcaacagaaaataaaaaggaatctcc	ttctccttggtctgaagagcc
Plate 3	G1	SC.t15.08	acaatcctatgtgccatcatcc	gaagttgccacagttaaccg
Plate 3	G2	SC.t15.09	gttaacattacccggtttttgaatc	tactctgttgtcttctttggaagg
Plate 3	G3	SC.t15.10	gggcgaaaaataatattctgcttcc	tatattactcccttgcgccac
Plate 3	G4	SC.t15.11	ttatttaaagttatgtcaccgtatgttagg	cccttaggcttttacatatatcatttgg

[Table continued on following page]

Plate	Well	tRNA ID	Fwd Primer	Reverse Primer
Plate 3	G5	SC.t15.12	ggaaaactaaaatgaaaaaaaataaattgggg	ttccgtttacctgcttcataatacc
Plate 3	G6	SC.t15.13	ccattttcttctactagggaagttcg	aacaattgttaaatctagtcaaaaagggg
Plate 3	G7	SC.t15.14	ggctattgaagtctcatcaggtatg	ggaaggaaaaacgtgtgcatc
Plate 3	G8	SC.t15.15	acaatctagaaagcgcagatagc	acactttctacggctatgcag
Plate 3	G9	SC.t15.16	cggagatcctaaaaatatgcttcg	taaagttatgcaaacgcctcac
Plate 3	G10	SC.t15.17	caatgaaactgtctgagatgatcag	gggatgaagaatatatagcagtgacg
Plate 3	G11	SC.t15.18	cctctcttatgggattaagaaatatcaac	cgtgatcagaaaaggaagatcatc
Plate 3	G12	SC.t15.19	catatataagttccacatgccatgag	tcgtacttatatggtatttgcagtcc
Plate 3	H1	SC.t15.20	gagtgtgcttctacagagcc	acctgtctagtaagattagcgc
Plate 3	H2	AmpR	ATGCTTAATCAGTGAGGCACCTATCTCAGC	GAATGAAGCCATACCAAACGACGAGCG

Appendix IV: Matching of tRNA Genes to Flanking Sequences

Table III: tRNA genes and their respective assigned flanking sequences. The tRNA matched column contains the technical tRNA gene names for *A. gossypii* (AGOS) and *E. cymbalariae* (Ecym). This table used information generated in part by programming scripts written by Mr. Isaac Luo.

tRNA ID	tRNA Name	Anticodon	Amino Acid	tRNA matched	Matching Anticodon	Matching Amino Acid
SC.t01.01	tP(UGG)A	TGG	Pro	AGOS_t0136	TGG	Pro
SC.t01.02	tA(UGC)A	TGC	Ala	AGOS_t0153	TGC	Ala
SC.t01.03	tL(CAA)A	CAA	Leu	AGOS_t0007	CAA	Leu
SC.t01.04	tS(AGA)A	AGA	Ser	AGOS_t0022	AGA	Ser
SC.t02.01	tL(UAA)B1	TAA	Leu	AGOS_t0116	TAA	Leu
SC.t02.02	tF(GAA)B	GAA	Phe	AGOS_t0006	GAA	Phe
SC.t02.03	tI(AAU)B	AAT	Ile	AGOS_t0062	AAT	Ile
SC.t02.04	tG(GCC)B	GCC	Gly	AGOS_t0002	GCC	Gly
SC.t02.05	tS(AGA)B	AGA	Ser	AGOS_t0043	AGA	Ser
SC.t02.06	tT(AGU)B	AGT	Thr	AGOS_t0102	AGT	Thr
SC.t02.07	tV(UAC)B	TAC	Val	AGOS_t0004	TAC	Val
SC.t02.08	tL(UAA)B2	TAA	Leu	Ecym_3229	TAA	Leu
SC.t02.09	tQ(UUG)B	TTG	Gln	AGOS_t0065	TTG	Gln
SC.t02.10	tR(UCU)B	TCT	Arg	AGOS_t0010	TCT	Arg
SC.t02.11	tD(GUC)B	GTC	Asp	AGOS_t0030	GTC	Asp
SC.t02.12	tC(GCA)B	GCA	Cys	AGOS_t0079	GCA	Cys
SC.t02.13	tE(UUC)B	TTC	Glu	AGOS_t0036	TTC	Glu
SC.t03.01	tE(UUC)C	TTC	Glu	AGOS_t0015	TTC	Glu
SC.t03.02	tL(CAA)C	CAA	Leu	AGOS_t0001	CAA	Leu
SC.t03.03	tP(AGG)C	AGG	Pro	AGOS_t0011	AGG	Pro
SC.t03.04	tN(GUU)C	GTT	Asn	AGOS_t0018	GTT	Asn
SC.t03.05	tG(GCC)C	GCC	Gly	AGOS_t0008	GCC	Gly
SC.t03.06	tM(CAU)C	CAT	Met	AGOS_t0035	CAT	Met
SC.t03.07	tK(CUU)C	CTT	Lys	AGOS_t0028	CTT	Lys
SC.t03.08	tQ(UUG)C	TTG	Gln	AGOS_t0003	TTG	Gln
SC.t03.09	tS(CGA)C	CGA	Ser	AGOS_t0017	CGA	Ser
SC.t03.10	tT(AGU)C	AGT	Thr	AGOS_t0073	AGT	Thr
SC.t04.01	tG(GCC)D1	GCC	Gly	AGOS_t0054	GCC	Gly
SC.t04.02	tK(UUU)D	TTT	Lys	AGOS_t0020	TTT	Lys
SC.t04.03	tA(AGC)D	AGC	Ala	AGOS_t0014	AGC	Ala
SC.t04.04	tT(AGU)D	AGT	Thr	AGOS_t0114	AGT	Thr
SC.t04.05	tS(AGA)D1	AGA	Ser	AGOS_t0082	AGA	Ser
SC.t04.06	tV(UAC)D	TAC	Val	AGOS_t0059	CAC	Val
SC.t04.07	tL(UAA)D	TAA	Leu	Ecym_4426	GAG	Leu

[Table continued on following page]

tRNA ID	tRNA Name	Anticodon	Amino Acid	tRNA matched	Matching Anticodon	Matching Amino Acid
SC.t04.08	tQ(UUG)D1	TTG	Gln	AGOS_t0121	TTG	Gln
SC.t04.09	tR(UCU)D	TCT	Arg	AGOS_t0053	TCT	Arg
SC.t04.10	tD(GUC)D	GTC	Asp	AGOS_t0045	GTC	Asp
SC.t04.11	tR(ACG)D	ACG	Arg	AGOS_t0051	ACG	Arg
SC.t04.12	tQ(UUG)D2	TTG	Gln	AGOS_t0156	TTG	Gln
SC.t04.13	tI(AAU)D	AAT	Ile	AGOS_t0063	AAT	Ile
SC.t04.14	tQ(UUG)D3	TTG	Gln	Ecym_1352	TTG	Gln
SC.t04.15	tI(UAU)D	TAT	Ile	AGOS_t0178	TAT	Ile
SC.t04.16	tY(GUA)D	GTA	Tyr	AGOS_t0111	GTA	Tyr
SC.t04.17	tS(AGA)D2	AGA	Ser	AGOS_t0084	AGA	Ser
SC.t04.18	tG(GCC)D2	GCC	Gly	AGOS_t0094	GCC	Gly
SC.t04.19	tE(CUC)D	CTC	Glu	AGOS_t0025	CTC	Glu
SC.t04.20	tV(CAC)D	CAC	Val	AGOS_t0005	CAC	Val
SC.t04.21	tF(GAA)D	GAA	Phe	AGOS_t0077	GAA	Phe
SC.t04.22	tX(XXX)D	???	Undet	Ecym_6174	???	Leu
SC.t04.23	tM(CAU)D	CAT	Met	AGOS_t0052	CAT	Met
SC.t04.24	tK(CUU)D1	CTT	Lys	AGOS_t0032	CTT	Lys
SC.t04.25	tG(CCC)D	CCC	Gly	AGOS_t0060	CCC	Gly
SC.t04.26	tS(AGA)D3	AGA	Ser	AGOS_t0177	AGA	Ser
SC.t04.27	tK(CUU)D2	CTT	Lys	AGOS_t0034	CTT	Lys
SC.t04.28	tL(CAA)D	CAA	Leu	AGOS_t0091	CAA	Leu
SC.t05.01	tG(GCC)E	GCC	Gly	AGOS_t0095	GCC	Gly
SC.t05.02	tS(AGA)E	AGA	Ser	Ecym_3557	AGA	Ser
SC.t05.03	tM(CAU)E	CAT	Met	AGOS_t0078	CAT	Met
SC.t05.04	tQ(UUG)E2	TTG	Gln	Ecym_2708	TTG	Gln
SC.t05.05	tK(CUU)E1	CTT	Lys	AGOS_t0066	CTT	Lys
SC.t05.06	tR(UCU)E	TCT	Arg	AGOS_t0072	TCT	Arg
SC.t05.07	tE(UUC)E1	TTC	Glu	AGOS_t0186	TTC	Glu
SC.t05.08	tH(GUG)E1	GTG	His	AGOS_t0021	GTG	His
SC.t05.09	tQ(UUG)E1	TTG	Gln	Ecym_4374	TTG	Gln
SC.t05.10	tS(UGA)E	TGA	Ser	AGOS_t0067	TGA	Ser
SC.t05.11	tA(UGC)E	TGC	Ala	AGOS_t0166	TGC	Ala
SC.t05.12	tE(UUC)E2	TTC	Glu	Ecym_3090	TTC	Glu
SC.t05.13	tH(GUG)E2	GTG	His	AGOS_t0026	GTG	His
SC.t05.14	tK(CUU)E2	CTT	Lys	AGOS_t0087	CTT	Lys
SC.t05.15	tV(AAC)E1	AAC	Val	AGOS_t0023	AAC	Val
SC.t05.16	tI(AAU)E1	AAT	Ile	AGOS_t0099	AAT	Ile
SC.t05.17	tV(AAC)E2	AAC	Val	AGOS_t0033	AAC	Val
SC.t05.18	tE(UUC)E3	TTC	Glu	Ecym_4588	TTC	Glu
SC.t05.19	tR(ACG)E	ACG	Arg	AGOS_t0176	ACG	Arg
SC.t05.20	tI(AAU)E2	AAT	Ile	AGOS_t0185	AAT	Ile
SC.t06.01	tP(UGG)F	TGG	Pro	AGOS_t0075	TGG	Pro
SC.t06.02	tN(GUU)F	GTT	Asn	AGOS_t0048	GTT	Asn
SC.t06.03	tF(GAA)F	GAA	Phe	AGOS_t0171	GAA	Phe

[Table continued on following page]

tRNA ID	tRNA Name	Anticodon	Amino Acid	tRNA matched	Matching Anticodon	Matching Amino Acid
SC.t06.04	tG(GCC)F1	GCC	Gly	AGOS_t0012	GCC	Gly
SC.t06.05	tY(GUA)F1	GTA	Tyr	AGOS_t0029	GTA	Tyr
SC.t06.06	tG(GCC)F2	GCC	Gly	AGOS_t0046	GCC	Gly
SC.t06.07	tS(GCU)F	GCT	Ser	AGOS_t0105	GCT	Ser
SC.t06.08	tA(AGC)F	AGC	Ala	AGOS_t0154	AGC	Ala
SC.t06.09	tY(GUA)F2	GTA	Tyr	AGOS_t0130	GTA	Tyr
SC.t06.10	tK(CUU)F	CTT	Lys	AGOS_t0109	CTT	Lys
SC.t07.01	tV(AAC)G3	AAC	Val	AGOS_t0037	AAC	Val
SC.t07.02	tH(GUG)G1	GTG	His	AGOS_t0122	GTG	His
SC.t07.03	tK(UUU)G1	TTT	Lys	AGOS_t0167	TTT	Lys
SC.t07.04	tK(CUU)G1	CTT	Lys	AGOS_t0120	CTT	Lys
SC.t07.05	tK(CUU)G2	CTT	Lys	AGOS_t0193	CTT	Lys
SC.t07.06	tL(CAA)G1	CAA	Leu	AGOS_t0138	CAA	Leu
SC.t07.07	tW(CCA)G1	CCA	Trp	AGOS_t0081	CCA	Trp
SC.t07.08	tH(GUG)G2	GTG	His	AGOS_t0151	GTG	His
SC.t07.09	tE(UUC)G1	TTC	Glu	Ecym_7052	TTC	Glu
SC.t07.10	tE(UUC)G2	TTC	Glu	AGOS_t0041	CTC	Glu
SC.t07.11	tR(UCU)G1	TCT	Arg	AGOS_t0118	TCT	Arg
SC.t07.12	tV(AAC)G1	AAC	Val	AGOS_t0058	AAC	Val
SC.t07.13	tL(CAA)G2	CAA	Leu	AGOS_t0144	CAA	Leu
SC.t07.14	tF(GAA)G	GAA	Phe	AGOS_t0117	GAA	Phe
SC.t07.15	tD(GUC)G1	GTC	Asp	AGOS_t0071	GTC	Asp
SC.t07.16	tE(UUC)G3	TTC	Glu	AGOS_t0148	CTC	Glu
SC.t07.17	tD(GUC)G2	GTC	Asp	AGOS_t0147	GTC	Asp
SC.t07.18	tS(AGA)G	AGA	Ser	Ecym_5450	AGA	Ser
SC.t07.19	tT(UGU)G1	TGT	Thr	AGOS_t0076	TGT	Thr
SC.t07.20	tL(GAG)G	GAG	Leu	AGOS_t0019	GAG	Leu
SC.t07.21	tK(UUU)G2	TTT	Lys	AGOS_t0170	TTT	Lys
SC.t07.22	tC(GCA)G	GCA	Cys	AGOS_t0104	GCA	Cys
SC.t07.23	tN(GUU)G	GTT	Asn	AGOS_t0038	GTT	Asn
SC.t07.24	tR(UCU)G3	TCT	Arg	AGOS_t0172	TCT	Arg
SC.t07.25	tI(AAU)G	AAT	Ile	AGOS_t0162	AAT	Ile
SC.t07.26	tA(AGC)G	AGC	Ala	AGOS_t0031	AGC	Ala
SC.t07.27	tG(UCC)G	TCC	Gly	AGOS_t0056	TCC	Gly
SC.t07.28	tA(UGC)G	TGC	Ala	AGOS_t0179	TGC	Ala
SC.t07.29	tV(AAC)G2	AAC	Val	AGOS_t0089	AAC	Val
SC.t07.30	tR(UCU)G2	TCT	Arg	AGOS_t0180	TCT	Arg
SC.t07.31	tG(GCC)G1	GCC	Gly	AGOS_t0110	GCC	Gly
SC.t07.32	tL(CAA)G3	CAA	Leu	AGOS_t0150	CAA	Leu
SC.t07.33	tK(CUU)G3	CTT	Lys	Ecym_2422	CTT	Lys
SC.t07.34	tW(CCA)G2	CCA	Trp	AGOS_t0133	CCA	Trp
SC.t07.35	tG(GCC)G2	GCC	Gly	AGOS_t0131	GCC	Gly
SC.t07.36	tT(UGU)G2	TGT	Thr	Ecym_4453	TGT	Thr
SC.t08.01	tH(GUG)H	GTG	His	Ecym_1316	GTG	His

[Table continued on following page]

tRNA ID	tRNA Name	Anticodon	Amino Acid	tRNA matched	Matching Anticodon	Matching Amino Acid
SC.t08.02	tV(AAC)H	AAC	Val	AGOS_t0127a	AAC	Val
SC.t08.03	tT(AGU)H	AGT	Thr	AGOS_t0125	AGT	Thr
SC.t08.04	tS(AGA)H	AGA	Ser	Ecym_7359	AGA	Ser
SC.t08.05	tQ(UUG)H	TTG	Gln	Ecym_4746	TTG	Gln
SC.t08.06	tA(AGC)H	AGC	Ala	AGOS_t0064	AGC	Ala
SC.t08.07	tF(GAA)H1	GAA	Phe	AGOS_t0194	GAA	Phe
SC.t08.08	tF(GAA)H2	GAA	Phe	Ecym_2291	GAA	Phe
SC.t08.09	tP(UGG)H	TGG	Pro	AGOS_t0137	TGG	Pro
SC.t08.10	tT(UGU)H	TGT	Thr	AGOS_t0157	CGA	Ser
SC.t08.11	tV(CAC)H	CAC	Val	AGOS_t0047	CAC	Val
SC.t09.01	tT(AGU)I1	AGT	Thr	AGOS_t0070	AGT	Thr
SC.t09.02	tI(AAU)I1	AAT	Ile	AGOS_t0016	AAT	Ile
SC.t09.03	tE(CUC)I	CTC	Glu	AGOS_t0024	CTC	Glu
SC.t09.04	tI(AAU)I2	AAT	Ile	AGOS_t0050	AAT	Ile
SC.t09.05	tS(UGA)I	TGA	Ser	AGOS_t0061	TGA	Ser
SC.t09.06	tK(CUU)I	CTT	Lys	Ecym_2477	CTT	Lys
SC.t09.07	tD(GUC)I1	GTC	Asp	AGOS_t0009	GTC	Asp
SC.t09.08	tT(AGU)I2	AGT	Thr	AGOS_t0155	AGT	Thr
SC.t09.09	tD(GUC)I2	GTC	Asp	AGOS_t0027	GTC	Asp
SC.t09.10	tE(UUC)I	TTC	Glu	AGOS_t0042	CTC	Glu
SC.t10.01	tT(AGU)J	AGT	Thr	AGOS_t0190	AGT	Thr
SC.t10.02	tE(UUC)J	TTC	Glu	Ecym_1267	CTC	Glu
SC.t10.03	tA(AGC)J	AGC	Ala	AGOS_t0085	AGC	Ala
SC.t10.04	tD(GUC)J1	GTC	Asp	AGOS_t0181	GTC	Asp
SC.t10.05	tR(ACG)J	ACG	Arg	Ecym_3026	ACG	Arg
SC.t10.06	tY(GUA)J1	GTA	Tyr	AGOS_t0195	GTA	Tyr
SC.t10.07	tR(UCU)J1	TCT	Arg	Ecym_7279	TCT	Arg
SC.t10.08	tD(GUC)J2	GTC	Asp	Ecym_5129	GTC	Asp
SC.t10.09	tD(GUC)J3	GTC	Asp	Ecym_7285	GTC	Asp
SC.t10.10	tR(UCU)J2	TCT	Arg	Ecym_7430	TCT	Arg
SC.t10.11	tV(AAC)J	AAC	Val	AGOS_t9173	AAC	Val
SC.t10.12	tM(CAU)J1	CAT	Met	AGOS_t0083	CAT	Met
SC.t10.13	tG(GCC)J1	GCC	Gly	AGOS_t0163	GCC	Gly
SC.t10.14	tK(CUU)J	CTT	Lys	Ecym_4352	CTT	Lys
SC.t10.15	tW(CCA)J	CCA	Trp	AGOS_t0143	CCA	Trp
SC.t10.16	tM(CAU)J2	CAT	Met	AGOS_t0088	CAT	Met
SC.t10.17	tL(UAA)J	TAA	Leu	AGOS_t9174	TAG	Leu
SC.t10.18	tM(CAU)J3	CAT	Met	AGOS_t0107	CAT	Met
SC.t10.19	tS(AGA)J	AGA	Ser	AGOS_t0165	CGC	Ala
SC.t10.20	tG(GCC)J2	GCC	Gly	Ecym_1335	GCC	Gly
SC.t10.21	tR(CCU)J	CCT	Arg	AGOS_t0158	CCT	Arg
SC.t10.22	tD(GUC)J4	GTC	Asp	AGOS_t0168	CGC	Ala
SC.t10.23	tY(GUA)J2	GTA	Tyr	AGOS_t0188	GTA	Tyr
SC.t10.24	tL(UAG)J	TAG	Leu	AGOS_t0057	TAG	Leu

[Table continued on following page]

tRNA ID	tRNA Name	Anticodon	Amino Acid	tRNA matched	Matching Anticodon	Matching Amino Acid
SC.t11.01	tT(CGU)K	CGT	Thr	Ecym_3610	CGT	Thr
SC.t11.02	tN(GUU)K	GTT	Asn	AGOS_t0103	GTT	Asn
SC.t11.03	tL(UAA)K	TAA	Leu	AGOS_t0129b	TAG	Leu
SC.t11.04	tE(UUC)K	TTC	Glu	Ecym_2245	CTC	Glu
SC.t11.05	tR(UCU)K	TCT	Arg	Ecym_8238	TCT	Arg
SC.t11.06	tK(CUU)K	CTT	Lys	Ecym_5431	CTT	Lys
SC.t11.07	tA(AGC)K1	AGC	Ala	AGOS_t0112	AGC	Ala
SC.t11.08	tW(CCA)K	CCA	Trp	AGOS_t0192	CCA	Trp
SC.t11.09	tV(AAC)K1	AAC	Val	AGOS_t0189	AAC	Val
SC.t11.10	tH(GUG)K	GTG	His	Ecym_4431	GTG	His
SC.t11.11	tV(AAC)K2	AAC	Val	AGOS_t0127b	AAC	Val
SC.t11.12	tL(CAA)K	CAA	Leu	Ecym_1315	CAA	Leu
SC.t11.13	tR(ACG)K	ACG	Arg	Ecym_3582	ACG	Arg
SC.t11.14	tD(GUC)K	GTC	Asp	Ecym_2339	GCT	Ser
SC.t11.15	tA(AGC)K2	AGC	Ala	AGOS_t0115	AGC	Ala
SC.t11.16	tK(UUU)K	TTT	Lys	Ecym_2779	TTT	Lys
SC.t12.01	tP(UGG)L	TGG	Pro	AGOS_t0152	TGG	Pro
SC.t12.02	tS(AGA)L	AGA	Ser	AGOS_t0055	CGA	Ser
SC.t12.03	tA(UGC)L	TGC	Ala	Ecym_3501	TGC	Ala
SC.t12.04	tR(ACG)L	ACG	Arg	Ecym_4237	ACG	Arg
SC.t12.05	tD(GUC)L1	GTC	Asp	Ecym_2487	CGA	Ser
SC.t12.06	tQ(UUG)L	TTG	Gln	AGOS_t0126	CTG	Gln
SC.t12.07	tL(UAG)L1	TAG	Leu	AGOS_t0080	TAG	Leu
SC.t12.08	tI(UAU)L	TAT	Ile	Ecym_4148	TAT	Ile
SC.t12.09	tL(CAA)L	CAA	Leu	Ecym_2338	CAA	Leu
SC.t12.10	tA(AGC)L	AGC	Ala	Ecym_1408	AGC	Ala
SC.t12.11	tV(AAC)L	AAC	Val	Ecym_3045	AAC	Val
SC.t12.12	tL(UAG)L2	TAG	Leu	AGOS_t0129a	TAG	Leu
SC.t12.13	tI(AAU)L1	AAT	Ile	AGOS_t0174	AAT	Ile
SC.t12.14	tX(XXX)L	GCT	Ser	AGOS_t0119	GCT	Ser
SC.t12.15	tD(GUC)L2	GTC	Asp	Ecym_2734	GCT	Ser
SC.t12.16	tE(UUC)L	TTC	Glu	AGOS_t0100	CTC	Glu
SC.t12.17	tR(CCG)L	CCG	Arg	AGOS_t0044	CCG	Arg
SC.t12.18	tK(UUU)L	TTT	Lys	Ecym_4428	TTT	Lys
SC.t12.19	tL(UAA)L	TAA	Leu	Ecym_6313	TAG	Leu
SC.t12.20	tN(GUU)L	GTT	Asn	AGOS_t0106	GTT	Asn
SC.t12.21	tI(AAU)L2	AAT	Ile	Ecym_2534	AAT	Ile
SC.t13.01	tR(UCU)M2	TCT	Arg	AGOS_t0164	CCG	Arg
SC.t13.02	tY(GUA)M1	GTA	Tyr	Ecym_2719	GTA	Tyr
SC.t13.03	tG(GCC)M	GCC	Gly	Ecym_2713	GCC	Gly
SC.t13.04	tP(UGG)M	TGG	Pro	AGOS_t0173	TGG	Pro
SC.t13.05	tS(AGA)M	AGA	Ser	Ecym_3050	CCA	Trp
SC.t13.06	tE(UUC)M	TTC	Glu	AGOS_t0113	CTC	Glu
SC.t13.07	tA(AGC)M1	AGC	Ala	Ecym_2798	AGC	Ala

[Table continued on following page]

tRNA ID	tRNA Name	Anticodon	Amino Acid	tRNA matched	Matching Anticodon	Matching Amino Acid
SC.t13.08	tF(GAA)M	GAA	Phe	Ecym_3227	GAA	Phe
SC.t13.09	tH(GUG)M	GTG	His	Ecym_3539	TCC	Gly
SC.t13.10	tV(AAC)M1	AAC	Val	Ecym_3120	AAC	Val
SC.t13.11	tW(CCA)M	CCA	Trp	Ecym_2284	CCA	Trp
SC.t13.12	tV(AAC)M2	AAC	Val	Ecym_5448	AAC	Val
SC.t13.13	tD(GUC)M	GTC	Asp	Ecym_3581	CTG	Gln
SC.t13.14	tK(CUU)M	CTT	Lys	Ecym_6022	CTT	Lys
SC.t13.15	tL(CAA)M	CAA	Leu	Ecym_3017	CAA	Leu
SC.t13.16	tM(CAU)M	CAT	Met	AGOS_t0108	CAT	Met
SC.t13.17	tV(AAC)M3	AAC	Val	AGOS_t0146	CAC	Val
SC.t13.18	tR(UCU)M1	TCT	Arg	Ecym_4327	TGA	Ser
SC.t13.19	tA(AGC)M2	AGC	Ala	Ecym_5235	AGC	Ala
SC.t13.20	tQ(CUG)M	CTG	Gln	AGOS_t0098	CTG	Gln
SC.t13.21	tY(GUA)M2	GTA	Tyr	Ecym_3164	GTA	Tyr
SC.t14.01	tG(UCC)N	TCC	Gly	AGOS_t0101	TCC	Gly
SC.t14.02	tN(GUU)N1	GTT	Asn	AGOS_t0175	GTT	Asn
SC.t14.03	tT(AGU)N1	AGT	Thr	Ecym_3116	AGT	Thr
SC.t14.04	tF(GAA)N	GAA	Phe	Ecym_7119	GAA	Phe
SC.t14.05	tL(CAA)N	CAA	Leu	Ecym_6161	CAA	Leu
SC.t14.06	tD(GUC)N	GTC	Asp	Ecym_4556	CCC	Gly
SC.t14.07	tP(UGG)N1	TGG	Pro	AGOS_t0187	TGG	Pro
SC.t14.08	tT(AGU)N2	AGT	Thr	Ecym_3601	AGT	Thr
SC.t14.09	tP(UGG)N2	TGG	Pro	Ecym_3141	TGG	Pro
SC.t14.10	tI(AAU)N1	AAT	Ile	Ecym_3322	AAT	Ile
SC.t14.11	tI(AAU)N2	AAT	Ile	Ecym_3502	AAT	Ile
SC.t14.12	tP(AGG)N	AGG	Pro	Ecym_1281	AGG	Pro
SC.t14.13	tN(GUU)N2	GTT	Asn	AGOS_t0182	GTT	Asn
SC.t14.14	tL(UAA)N	TAA	Leu	Ecym_6321	CAA	Leu
SC.t15.01	tG(UCC)O	TCC	Gly	Ecym_1515	TCC	Gly
SC.t15.02	tT(AGU)O1	AGT	Thr	Ecym_5372	AGT	Thr
SC.t15.03	tG(GCC)O1	GCC	Gly	Ecym_5641	GCC	Gly
SC.t15.04	tN(GUU)O1	GTT	Asn	Ecym_2687	GTT	Asn
SC.t15.05	tS(GCU)O	GCT	Ser	AGOS_t0149	GCT	Ser
SC.t15.06	tG(GCC)O2	GCC	Gly	Ecym_6019	GCC	Gly
SC.t15.07	tY(GUA)O	GTA	Tyr	Ecym_4622	GTA	Tyr
SC.t15.08	tP(UGG)O1	TGG	Pro	Ecym_4457	TGG	Pro
SC.t15.09	tR(ACG)O	ACG	Arg	Ecym_4627	GTA	Tyr
SC.t15.10	tT(AGU)O2	AGT	Thr	Ecym_5050	CAT	Met
SC.t15.11	tK(UUU)O	TTT	Lys	Ecym_8146	TTT	Lys
SC.t15.12	tP(UGG)O2	TGG	Pro	Ecym_4603	TGG	Pro
SC.t15.13	tN(GUU)O2	GTT	Asn	Ecym_3543	GTT	Asn
SC.t15.14	tD(GUC)O	GTC	Asp	Ecym_5220	GTA	Tyr
SC.t15.15	tG(CCC)O	CCC	Gly	AGOS_t0159	CCC	Gly
SC.t15.16	tV(AAC)O	AAC	Val	Ecym_3206	CAC	Val

[Table continued on following page]

tRNA ID	tRNA Name	Anticodon	Amino Acid	tRNA matched	Matching Anticodon	Matching Amino Acid
SC.t15.17	tM(CAU)O1	CAT	Met	Ecym_1381	CAT	Met
SC.t15.18	tA(UGC)O	TGC	Ala	Ecym_6070	TGC	Ala
SC.t15.19	tM(CAU)O2	CAT	Met	Ecym_4463	CAT	Met
SC.t15.20	tP(UGG)O3	TGG	Pro	Ecym_5673	TGG	Pro
SC.t16.01	tW(CCA)P	CCA	Trp	Ecym_2722	CCA	Trp
SC.t16.02	tE(UUC)P	TTC	Glu	AGOS_t0132	CTC	Glu
SC.t16.03	tM(CAU)P	CAT	Met	Ecym_4599	CAT	Met
SC.t16.04	tC(GCA)P1	GCA	Cys	AGOS_t0169	GCA	Cys
SC.t16.05	tF(GAA)P1	GAA	Phe	Ecym_7336	GAA	Phe
SC.t16.06	tG(GCC)P1	GCC	Gly	Ecym_6114	GCC	Gly
SC.t16.07	tK(CUU)P	CTT	Lys	Ecym_7214	CTT	Lys
SC.t16.08	tF(GAA)P2	GAA	Phe	Ecym_8186	GAA	Phe
SC.t16.09	tS(UGA)P	TGA	Ser	Ecym_3462	TGA	Ser
SC.t16.10	tT(UGU)P	TGT	Thr	Ecym_5366	GTT	Asn
SC.t16.11	tK(UUU)P	TTT	Lys	Ecym_5433	CCA	Trp
SC.t16.12	tC(GCA)P2	GCA	Cys	Ecym_5164	GCA	Cys
SC.t16.13	tN(GUU)P	GTT	Asn	Ecym_4412	GTT	Asn
SC.t16.14	tI(AAU)P1	AAT	Ile	Ecym_8367	AAT	Ile
SC.t16.15	tA(AGC)P	AGC	Ala	Ecym_7417	AGC	Ala
SC.t16.16	tG(GCC)P2	GCC	Gly	Ecym_6124	GCC	Gly
SC.t16.17	tI(AAU)P2	AAT	Ile	Ecym_6020	TGG	Pro

Appendix V: Genotypes of Yeast Strains Used in this Study

Table IV: Genotypes of yeast strains used in this study.

Strain ID	Genotype	Comments
RWy007	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1*	BY4741
RWy099	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3)-LEU2*-Syn.tRNA VI inchworm vector]	BY4741 + Chr6 tRNA array inchworm vector. Isolate 1. Neochromosome start point
RWy101	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 sup61Δ::KanMX ho::*Syn.*SUP61-URA3*	BY4741 *sup61*Δ::*KanMX* + *SUP61* tRNA cassette integrated into the *HO* locus. Isolate 1. Verified using colony PCR.
RWy103	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 sup61Δ::KanMX* [pRS415(*URA3*)-Syn.*SUP61*]	BY4741 *sup61*Δ::*KanMX* + *SUP61* tRNA cassette in pRS416. Isolate 1
RWy105	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 trt2Δ::KanMX ho::*Syn.*TRT2-URA3*	BY4741 *TRT2*Δ::*KanMX* + *TRT2* tRNA cassette integrated into the *HO* locus. Isolate 1. Verified using colony PCR.
RWy107	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 trt2Δ::KanMX* [pRS416(*URA3)*-Syn.*TRT2]*	BY4741 *TRT2Δ::KanMX* + *TRT2* tRNA cassette in pRS416. Isolate 1
RWy109	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 trr4Δ::KanMX ho::*Syn.*TRR4-URA3*	BY4741 *TRR4Δ::KanMX* + *TRR4* tRNA cassette integrated into the HO locus. Isolate 1. Verified using colony PCR.
RWy111	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS416(*URA3)*-Syn.*SUP61-TRT2-TRR4*]	BY4741 *TRR4Δ::KanMX* + *TRR4* tRNA cassette in pRS416. Isolate 1
RWy121	*MATα leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::Syn.SUP61::ura3 synVI SYN-WT.PRE4 IXL-synIXR*	Triple synthetic chromosome (Syn3, Syn6, Syn9) isolate 1. Genotype verified. Note: contains a partial *URA3* remnant in *HO* locus!
RWy125	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3)*-Syn.tRNA VII]	Triple synthetic chromosome (Syn3, Syn6, Syn9) + Chr6 tRNA array inchworm vector. Isolate 1. Neochromosome start point
RWy141	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*URA3*-Syn.tRNA VI-Syn.tRNA V]	Inchworm round 1 candidate. BY4741. Isolate 9. Digest pattern correct.
RWy144	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*URA3*-Syn.tRNA VII-Syn.tRNA VI]	Inchworm round 1 candidate. Triple Syn. Isolate 18. Digest pattern correct
RWy160	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3)-LEU2*-Syn.tRNA-VI-V-IV]	Inchworming round 2. BY4741. Isolate 13. PCR tag verified.
RWy161	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV]	Inchworming round 2. Triple-Syn-Chr. Isolate 7. PCR tag verified
RWy162	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3)-URA3*-Syn.tRNA-VI-V-IV-III]	Inchworming round 3. BY4741. Isolate 8 (from re-streak). PCR tag verified.

[Table continued on next page]

Strain ID	Genotype	Comments
RWy163	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III]	Inchworming round 3. Triple-Syn-Chr. Isolate 6. PCR tag verified.
RWy167	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3)-LEU2*-Syn.tRNA-VI-V-IV-III-II]	Inchworming Round 4 - BY4741. Introducing Chr2 tRNA array. Isolate 2. PCR tag verified
RWy168	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II]	Inchworming Round 4 - Triple Synthetic Chromosome. Introducing Chr2 tRNA array. Isolate 2. PCR tag verified
RWy169	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3)-URA3*-Syn.tRNA-VI-V-IV-III-II-I]	Inchworming round 5 - BY4741. Introducing Chr1 tRNA array. Isolate 5. PCR Tag verified
RWy170	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I]	Inchworming round 5 - Triple-Synthetic-Chromosome. Isolate 2. PCR Tag verified
RWy171	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII]	Inchworming round 6 - BY4741. Introducing Chr7 tRNA array. Isolate 2. PCR Tag verified.
RWy172	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII]	Inchworming round 6 - Triple-Synthetic-Chromosome. Isolate 3. PCR Tag verified
RWy177	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI]	Inchworming round 7 Candidate - Triple-Synthetic-Chromosome. Isolate 3. Partial PCR Tags. Full PCR Tag verified
RWy180	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI]	Inchworming round 7 Candidate - BY4741. Isolate 27. Partial PCR Tags verified. Full PCR Tag verified
RWy182	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX]	Inchworming round 8 - BY4741. Introducing Chr9. Isolate 11. Verified using PCR tags.
RWy185	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX]	Inchworming round 8 - triple-synthetic-chromosome. Introducing Chr9. Isolate 10. Verified using PCR tags.
RWy198	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII]	Inchworming round 9 - BY4741. Introducing Chr8 array. isolate 10. Partial PCR Tags good. Verified with full PCR tags.
RWy202	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII]	Inchworming round 9 - triple-synthetic-chromosome. Isolate 24. Introducing Chr8 array. isolate 24. Verified with full PCR tags.
RWy203	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X]	Inchworming round 10 - BY4741. Introducing Chr10 tRNA array. Isolate 7. Partial PCR Tags good. Verified with full PCR tags.

[Table continued on next page]

Strain ID	Genotype	Comments
RWy206	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X]	Inchworming round 10 - triple-synthetic-chromosome. Isolate 1. Introducing Chr10 tRNA array. Partial PCR Tags good. Verified with full PCR tags.
RWy215	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI]	Inchworming round 11 - BY4741. Introducing Chr11 tRNA array. Isolate 1. Partial PCR Tags good. Verified with full PCR tags. Full genome sequencing verified.
RWy217	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI]	Inchworming round 11 - triple-synthetic-chromosome. Isolate 4. Introducing Chr11 tRNA array. Partial PCR Tags good. Verified with full PCR tags. Full genome sequencing verified.
RWy218	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII]	Inchworming round 12 - BY4741. Introducing Chr13 tRNA array. Isolate 32. Partial PCR Tags good. Verified with full PCR tags.
RWy220	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII]	Inchworming round 12 - triple-synthetic-chromosome. Isolate 1. Introducing Chr13 tRNA array. Partial PCR Tags good. Verified with full PCR tags.
RWy222	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 sup61Δ::KanMX ho::*Syn.*SUP61*(WT 5')-*URA3*	BY4741 *SUP61Δ::KanMX* + *SUP61* tRNA cassette containing a WT 5' flanking sequence integrated into the *HO* locus. Isolate 1A. Verified with colony PCR. *SUP61* cassette on pRS415 plasmid shuffled out.
RWy224	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 sup61Δ::KanMX ho::*Syn.*SUP61*(WT 3')-*URA3*	BY4741 *SUP61Δ::KanMX* + *SUP61* tRNA cassette containing a WT 3' flanking sequence integrated into the *HO* locus. Isolate 2A. Verified with colony PCR. SUP61 cassette on pRS415 plasmid shuffled out.
RWy226	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 sup61Δ::KanMX ho::*Syn.*SUP61*(Extra Bases)-*URA3*	BY4741 *SUP61Δ::KanMX* + *SUP61* tRNA cassette containing extra bases between PolyT integrated into the *HO* locus. Isolate 3A. Verified with colony PCR. *SUP61* cassette on pRS415 plasmid shuffled out.
RWy228	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 sup61Δ::KanMX ho::*Syn.*SUP61*(No Intron)-*URA3*	BY4741 *SUP61Δ::KanMX* + WT *SUP61* tRNA cassette with no intron integrated into the *HO* locus. Isolate 4A. Verified with colony PCR. *SUP61* cassette on pRS415 plasmid shuffled out.
RWy230	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 sup61Δ::KanMX ho::*WT.*SUP61-URA3*	BY4741 *SUP61Δ::KanMX* + WT *SUP61* tRNA cassette integrated into the *HO* locus. Isolate 5A. Verified with colony PCR. *SUP61* cassette on pRS415 plasmid shuffled out.

[Table continued on next page]

Strain ID	Genotype	Comments
RWy232	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII]	Inchworming round 13 - BY4741. Introducing Chr12 tRNA array. Isolate 24. Partial PCR Tags good. Verified with full PCR tags (note gel not the best).
RWy235	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII]	Inchworming round 13 - triple-synthetic-chromosome. Isolate 4. Introducing Chr12 tRNA array. Partial PCR Tags good.
RWy237	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV]	Inchworming Round 14 - BY4741. Introducing Chr14 tRNA Array. Isolate 8. Partial PCR Tags good. Verified with full PCR tags
RWy242	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*LEU2*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV]	Inchworming Round 14 - Triple-Synthetic-Chromosome. Introducing Chr14 tRNA Array. Isolate 20. Partial PCR Tags good. Verified with full PCR tags
RWy243	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV]	Complete circular neochromosome. Inchworming Round 15 - BY4741. Introducing Chr15 tRNA Array. Isolate 9. Partial PCR Tags good. Verified with full PCR tags. Deep sequencing not clear - apparently a plasmid duplication + an internal deletion.
RWy248	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV]	Complete circular neochromosome. Inchworming Round 15 - Triple-Synthetic-Chromosome. Introducing Chr15 tRNA Array. Isolate 24. Verified with full PCR tags. Deep sequencing reveals two mixed-copy plasmids with a 1.17 kb deletion.
RWy254	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV]	Neochromosome *URA3* pop-out. BY4741 isolate 14. Pop-out verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.
RWy256	MAT*α leu2Δ0 lys2Δ0 MET15 ura3Δ0 his3Δ1 synIII ho::*Syn.*SUP61::ura3 synVI SYN-WT.PRE4 IXL-synIXR* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV]	Neochromosome *URA3* pop-out. Triple-synthetic-chromosome isolate 6. Pop-out verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.
RWy262	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-ampR::Telomerator-*URA3*]	Neochromosome + telomerator integrated next to CEN (Cassette A). *URA3* previously popped-out. BY4741 isolate 1. Integration verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.

[Table continued on next page]

Strain ID	Genotype	Comments
RWy263	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter3Δ::Telomerator-*URA3*]	Neochromosome + telomerator replacing Ter3 (Cassette B). *URA3* previously popped-out. BY4741 isolate 2 (well A3). Integration verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.
RWy265	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter2Δ::Telomerator-*URA3*]	Neochromosome + telomerator replacing Ter2 (Cassette C). URA3 previously popped-out. BY4741 isolate 1 (well C4/B4). Integration verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.
RWy276	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [Neochromosome(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter3Δ::TeSS-*ura3*] [pRS415(*LEU2*)-I-SceI]	Linearised tRNA neochromosome. Cassette B isolate 2 (*Ter3* proximal). Note: still contains *I-SceI* on pRS415! Produces a 240 kb band (60 kb too big). Quick colony PCR good. Deep sequencing reveals a complex duplication.
RWy277	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [Neochromosome(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter3Δ::TeSS-*ura3*]	Linearised tRNA neochromosome. Cassette B isolate 6 (Ter3 proximal). Strain appears to have lost *I-SceI* on pRS415. Produces a 240 kb band (60kb too big). Quick colony PCR good
RWy278	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [Neochromosome(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter2Δ::TeSS-*ura3*]	Linearised tRNA neochromosome. Cassette C isolate 10 (Ter2 Proximal). Strain appears to have lost *I-SceI* on pRS415. Produces a 350 kb band (170kb too big!). Quick colony PCR good. Deep sequencing reveals a large duplication.
RWy279	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [Neochromosome(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter2Δ::TeSS-*ura3*]	Linearised tRNA neochromosome. Cassette B isolate 12 (Ter2 proximal). Strain appears to have lost *I-SceI* on pRS415. Produces a 260 kb band (80kb too big). Quick colony PCR good. Deep sequencing reveals a large duplication (similar to RWy278)
RWy280	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-*URA3*-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV]	Circular tRNA neochromosome (triple-synthetic-chromosome background) transferred to BY4741. Isolate B1. Quick colony PCR good. Not yet verified with full PCR tags.
RWy287	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter3Δ::Telomerator-*URA3*]	Neochromosome (triple-synthetic-chromosome background) + telomerator replacing Ter3 (Cassette B). *URA3* popped-out. BY4741 isolate A2. Integration verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.

[Table continued on next page]

Strain ID	Genotype	Comments
RWy290	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter2Δ::Telomerator-*URA3*]	Neochromosome (triple-synthetic-chromosome background) + telomerator replacing Ter2 (Cassette C). *URA3* popped-out. BY4741 isolate B2. Integration verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.
RWy301	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [Neochromosome(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter3Δ::TeSS-*ura3*] [pRS415(LEU2)-I-SceI]	Linearised tRNA neochromosome (triple-synthetic background). BY4741 cassette B (A2-1) isolate 2 (Ter3 proximal). Still contains I-SceI on pRS415. Quick colony PCR good. Produces a ~190kb PFGE band.
RWy303	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [Neochromosome(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV-Ter2Δ::TeSS-*ura3*]	Linearised tRNA neochromosome (triple-synthetic background). BY4741 cassette C (B2-1) isolate 10 (Ter2 proximal). Quick colony PCR good. Produces a ~190 kb PFGE band.
RWy307	MAT**a** *leu2Δ0 met15Δ0 ura3Δ0 his3Δ1* [pRS413(*HIS3*)-Syn.tRNA-VI-V-IV-III-II-I-VII-XVI-IX-VIII-X-XI-XIII-XII-XIV-XV]	Neochromosome URA3 pop-out. BY4741 isolate A10. Pop-out verified using colony PCR. Quick colony PCR tags good. Not yet verified with full PCR tags.

Appendix VI: PCR Primer Sequences Used in this Study

Table V: PCR primer sequences used in this study. The following primers were described in Chapter 4 and Chapter 5 of this thesis.

Primer ID	Direction	Sequence
YCp469	Fwd	ccgcaatactatgtagctaaact
YCp471	Rev	tgagtcaagacctcgacaacat
YCp1807	Rev	ATGAAATTCATAATAGAAACAATAACTAGCTGTACGACTTCAGCAATTAGCTACTCG
YCp1809	Fwd	AAGTCGTACAGCTAGTTATTGTTTCTATTATGAATTTCATTTATAAAGTTTATGTAC
YCp1812	Rev	GTCATCACCGAAACGCGCGAAATAACTAGCTGTACGACTTCAGCAATTAGCTACTCG
YCp1813	Fwd	GCTGAAGTCGTACAGCTAGTTATTTCGCGCGTTTCGGTGATGACGGTG
YCp2581	Fwd	TGAGATCCAGTTCGATGTAACCC
YCp2582	Rev	GCTGAAGATCAGTTGGGTGC
YCp2606	Fwd	GCCTTAGACCGCTCGGCCAAGCTTGATATCGAATTCCTGCAGC
YCp2607	Rev	TACGCCGAGCCCAGGATTGTGCGGCCGCGGCCGGCCTTATCGATACCGTCGACCTCGAG
YCp2608	Fwd	GAGGTCGACGGTATCGATAAGGCCGGCCGCGGCCGCACAATCCTGGGCTCGGCGT
YCp2609	Rev	GCAGGAATTCGATATCAAGCTTGGCCGAGCGGTCTAAGG
YCp2610	Fwd	aattagagcTTCAATTTAATGCTTGATATCGAATTCCTGCAGC
YCp2611	Rev	GCAGGAATTCGATATCAAGCATTAAATTGAAgctctaatttgtgagtttagtatacATGC
YCp2612	Fwd	GCACTCGATCTTCCCAGAAAAAGAGGC
YCp2613	Rev	TGTTCTGCTACTGCTTCTGCCTC
YCp2786	Fwd	gtggatgggttcgatttaactttaaataatgccaattatttaaagttaGCGGCCGCGCTTGATATCGAATTCCTGCAGC
YCp2787	Rev	gttttttccgtagatataactttaaataattggcattatttaaagttaGCGGCCGCTTATCGATACCGTCGACCTCGAG
YCp2937	Fwd	gagcctattggtccctgtgagagc
YCp2938	Rev	cgtctttttgaaatagccacaatgcaa
YCp2939	Fwd	ccttcaagcaacagtctacagtgtac
YCp2940	Rev	gagatctcctctcggatatcacagg
YCp2968	Fwd	agtaccttgcgcttcagatcttg
YCp2969	Rev	ccatccgccgctgatatcaag

[Table continued on next page]

Primer ID	Direction	Sequence
YCp3001 (M13F)	Fwd	GTAAAACGACGGCCAGT
YCp3002 (M13R)	Rev	CAGGAAACAGCTATGAC
YCp3084	N/A	ggtccgttgcggactcggatatgg
YCp3100	N/A	ctgcggtcaagatatttcttG
YCp3114	Fwd	TTTTTCACGGATTTTTTGTGCCG
YCp3115	Rev	CCCATGGTGATGCAGGAGCA
YCp3479	Fwd	cccaaatctctctatatcacgtgc
YCp3480	Rev	tactgcgcctttacaacatcg
YCp3722	Fwd	tttttgtgcctttagcatcacatg
YCp3723	Rev	gaaaagttactgctctaatcaacagg
YCp3724	Fwd	gcatatctttaccagaatatggtgatc
YCp3725	Rev	ctggacatagcgtattatgttgc
YCp3741	Fwd	acgtgacactgaaaataactttaaataattggcattatttaaagttaTTTTTATTTTTTTTTatcaactgtccgggcac
YCp3742	Rev	CGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGCCGGCCTTTTTTTTTCTTCCTCTTCATTTTTCATTGCAAAG
YCp3743	Fwd	AGACCGCTCGGCCAAGCTTGATATCGAATTCCTGCAGCCCtatataacagtatatgctgactttacgatggc
YCp3744	Rev	agaacaatgacgtatattaactttaaataatgccaattatttaaagttatatctacggaaaaaaccgccaaaaataaag
YCp4177	Fwd	TCGAGGTCGACGGTATCG
YCp4178	Rev	CGTTCTTCCTTCTGTTCGGAG
YCp4488	Rev	aatgacagcactacttgcg